DOCUMENT RESUME

ED 181 075 TM 010 026

AUTHOR	Rolph, John E.: And Others
TITLE	Predicting Minority and Majority Student Performance on the National Board Exams.
INSTITUTION	Rand Corp., Santa Monica, Calif.
SPONS AGENCY	Health Resources Administration (DHEW/PHS), Bethesda, Md. Div. of Medicine.
REPORT NO	Rand-R-2029-HEW
PUB DATE	Nov 78
CONTRACT	231-75-0614
NOTE	96p.
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	Admission Criteria: Affirmative Action: Average Students: College Choice: College Entrance Examinations: Grade Point Average: Higher Education: Institutional Characteristics: Medical Schools: *Medical Students: *Minority Groups: *Occupational Tests: Predictive Ability (Testing): *Predictive Measurement: Predictor Variables: *Racial Differences: Sex Differences: Standardized Tests: Technical Reports: Test Results
IDENTIFIERS	*Medical College Admission Test: *National Board of Medical Examiners Examination

ABSTRACT

A study was conducted to predict performance of majority and minority students in medical school by analyzing results of the National Board of Medical Examiners (NBME), Parts I and II. Data were collected from members of the 1975 and 1976 graduating classes of nine medical schools, and included Medical College Admission Test (MCAT) scores, undergraduate grade point averages (GPA), undergraduate school attended, sex, race, and medical school characteristics. Race was coded as Black, Asian, and other (mostly Mexican-Americans or Puerto Ricans) for minority students. It was concluded that: broad aptitude measures of the quantitative and verbal parts of the MCAT had greater predictive power for minority students than for majority students: measures of general aptitude were better predictors than measures of past achievement for minority students: the selectivity of the undergraduate college had a statistically significant positive effect on majority students' performance: the effects of medical school differences in relation to compensatory education programs were weak: and, for majority and minority medical students with equivalent premedical school characteristics, the majority students could be expected to score higher than the minority students on the NBME in two out of three cases. (Implications of this study were discussed). (MH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED181075

PREDICTING MINORITY AND MAJORITY MEDICAL STUDENT PERFORMANCE ON THE NATIONAL BOARD EXAMS

PREPARED FOR THE HEALTH RESOURCES ADMINISTRATION,
DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

JOHN E. ROLPH,
ALBERT P. WILLIAMS,
A. LEE LANIEAR
WITH THE ASSISTANCE OF WENDY D. COOPER

R-2029-HEW
NOVEMBER 1978

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

SCOPE OF INTEREST NOTICE
The ERIC Facility has assigned this document for processing to: TM UD
In our judgement, this document is also of interest to the clearinghouses noted to the right. Indexing should reflect their special points of view.

TM010 026

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Rand Corp.
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Rand
SANTA MONICA, CA 90406

The work upon which this publication is based was performed pursuant to Contract No. HRA-231-75-0614 with the Health Resources Administration, Department of Health, Education, and Welfare. Views or conclusions contained in this study should not be interpreted as representing the official opinion or policy of the Department of Health, Education, and Welfare.

LC Card No. 78-24782
ISBN 0-8330-0094-2

Published by The Rand Corporation
1700 Main Street, Santa Monica, Ca. 90406 Phone (213) 393-0411

Principal Investigator: Albert P. Williams

Contract Project Officer: Joseph Millard Brown
Division of Medicine
Public Health Services
Health Resources Administration
Department of Health, Education, and Welfare
Hyattsville, Maryland 20782

The work upon which this publication is based was performed pursuant to Contract No. HRA-231-75-0614 with the Health Resources Administration, Department of Health, Education, and Welfare. Views or conclusions contained this this study should not be interpreted as representing the official opinion or policy of the Department of Health, Education, and Welfare.

Published by The Rand Corporation
1700 Main Street, Santa Monica, Ca. 90406
Phone (213) 393-0411

Principal Investigator: Albert P. Williams

Contract Project Officer: Joseph Millard Brown
Division of Medicine, Public Health Services,
Health Resources Administration
Department of Health, Education, and Welfare
Hyattsville, Maryland 20782

PREFACE

The research in this report was prepared under contract with the Division of Medicine, Health Resources Administration, Department of Health, Education, and Welfare (Contract No. HRA-231-75-0614).

The primary objective of this research is to develop better means for predicting majority and minority student performance in medical schools and thus to assist admissions committees in evaluating medical school applicants with very different backgrounds. Data on the National Board of Medical Examiners exams are used to measure performance because these standardized tests are widely used by medical schools for student evaluation and have fairly unambiguous meaning across schools. This does not imply that these exams are the best criteria for evaluation.

This report should be of interest to those concerned with prediction of medical school performance for either minority or majority students. Since the data are from the classes of 1975 and 1976 of nine medical schools, the results should be of particular interest to those schools. The methodology used in the report may also be of interest to anyone concerned with applications of statistical data analysis.

This report is one of a series of three related studies dealing with medical school admissions and performance. The other two are:

John E. Rolph, Albert P. Williams, Carolyn L. Lee, *The Effect of State of Residence on Medical School Admissions*, R-2014-HEW, February 1978

Albert P. Williams, Wendy D. Cooper, Carolyn L. Lee, *Factors Affecting Medical School Admission Decisions for Minority and Non-Minority Applicants: A Comparative Study of Ten Schools*, R-2030-HEW, forthcoming

SUMMARY

This report deals with the problem of predicting performance of majority and minority students in medical school. The principal purposes are to aid admissions committees in evaluating medical school applicants with very different backgrounds and to assist third parties in assessing the effectiveness and efficiency of affirmative action programs.

U.S. medical schools faced new problems in predicting student performance as they began affirmative action programs targeted at minority groups. By the mid-1970s, minority recruitment and admissions had become a major concern of medical schools throughout the country. There is intense competition for the academically highly qualified minority applicant, and many school resources are devoted to screening minority applicants to find those who are able and motivated to complete the demanding medical school curriculum.

There is considerable debate over how, when, and by what standards it is appropriate to evaluate majority and minority medical student performance. We use scores on the National Board of Medical Examiners (NBME) Part I and Part II exams in our analysis, but, in doing so, we do not wish to enter the debate on the appropriateness of different performance standards. Rather we use these data because they are standardized measures, widely used by medical schools to evaluate individual students as well as programs.

In our prediction models, we use data on members of the graduating classes of 1975 and 1976 from nine medical schools. The data are similar to those available to every medical school when they decide which applicants to accept for admission: Medical College Admission Test (MCAT) scores, undergraduate grade point averages (GPA), and other descriptive data that we have quantified in ways that make them usable in statistical analyses. We recognize that we have not used other information (e.g., interviews, letters of recommendation) used by admission committees to predict student performance, but we have restricted our analysis to data that can be

interpreted unambiguously for any medical school.

It is unreasonable to expect minority students (mostly black, Asian and Mexican American in this study), whose preadmission characteristics (MCAT, GPA) at admission are on average well below the average majority student's characteristics, to have National Board scores that are comparable on average to majority students' scores. An important aspect of affirmative action programs is that admissions committees use different criteria for admitting minority students and majority students. Therefore, we develop one set of equations that predict majority student scores on the National Boards as a function of their preadmission characteristics and a different set for minority students. We also provide means for comparing majority and minority student equations and hence their performances with the same preadmission characteristics.

Our analysis was based on a much larger student population than is available for analysis by any single school, and that larger population will support a considerably more rigorous statistical analysis than can be performed using data from only one school. Although in many respects our prediction equations resemble those developed by others, our larger sample size gives a clearer understanding of the various factors related to performance than is possible for analysis of medical student data from a single school.

This expectation is realized in the majority analysis. For example, every predictor variable we used has statistically significant and substantial predictive power for at least some part of the NBME exams. Among these predictor variables was the general information part of the MCAT, which has recently been eliminated from the revised test in part because it was widely regarded as having minimal predictive power for performance in the medical school curriculum.

In analyzing mean Part I, biochemistry Part I, mean Part II, medicine Part II, surgery Part II, and psychiatry Part II of the NBME exam, we observe that there are quite substantial differences in the prediction models for the various parts of the NBME exam. This raises an issue regarding the appropriateness of tradeoffs in one area of performance for another, and logically that issue would be resolved

differently for specialist-oriented versus generalist-oriented medicine.

There are several notable differences between the minority and majority prediction models. This implies that medical school administrators will make unnecessarily large errors in predicting minority student performance if they extrapolate solely from their experience with majority students. In particular, broad aptitude measures of the quantitative and verbal parts of the MCAT exam have considerably more predictive power for minority than for majority students. Although the science MCAT and undergraduate GPA are the best predictors of majority students' performance in all but one of our models, verbal MCAT has better predictive power than either in all but one of the minority models. The quantitative MCAT is superior to the science MCAT in one of the minority models and superior to GPA in three. This suggests that measures of general aptitude may be better predictors of performance than measures of past achievement, within the ranges observed for minority medical students.

Another notable difference between minority and majority students is the effect of the undergraduate school attended. The selectivity of the undergraduate school has a statistically significant positive effect in majority performance models for four of our six parts of the NBME exam. That variable never approaches a level of statistical (or practical) significance in any of the six minority models.

An important goal of affirmative action programs is to discover ways to compensate for the generally inferior academic preparation of minority students. We attempted to identify medical schools that achieved unusual success with these compensatory programs, but we were unable to do so. In general, the effects of school differences on minority performance were weak. There is some evidence in the equations that school (or curriculum) effects on majority students are sometimes amplified for minority students. However, we are inclined not to view this as an indication of affirmative action program success or failure but rather as an indication that minority students have more to gain or lose in terms of better scores when a school emphasizes or de-emphasizes the NBME as a means of student evaluation.

We believe it is important to the effectiveness of affirmative action programs to empirically establish the different factors that predict majority and minority student performance. An important implication of our research is that it is possible to predict minority performance with as much or greater accuracy as majority performance. Currently, there is considerable debate regarding appropriate standards for affirmative action programs and who should set those standards. That debate should not proceed on the assumption that there are better statistical bases for predicting majority than minority performance on a standardized test such as the NBME exam.

The final part of our analysis examined differences between minority and majority student performance in several ways. The figures in Sec. IV show the distribution of actual performance of individual students from both groups about the prediction line. These figures show not only the spread of performance of the two groups but also provide a graphic picture of prediction accuracy. The reader can clearly see that the overlap in performance of the two groups is substantial and that although the average minority scores are lower, most members of both groups have scores in the same range.

We also compared the predicted performance of majority and minority students with equivalent characteristics. That analysis showed that, on *average*, a majority student's score is at least 22 points higher on mean Part I and at least 36 points higher on mean Part II than a minority student with the same characteristics. A consequence of this is that the odds are 3 to 2 or better that an *individual* majority student would score higher than a minority student with the same preadmission characteristics. We have no definitive explanation for that finding, but it is too consistent across models and student types to be dismissed as a statistical artifact.

CONTENTS

PREFACE iii

SUMMARY v

LIST OF TABLES xi

LIST OF FIGURES xiii

ACKNOWLEDGMENTS xv

Section

I. INTRODUCTION 1

II. METHODOLOGY 10

The Data 10
Statistical Modeling 12

III. ANALYSIS OF MAJORITY AND MINORITY STUDENT PERFORMANCE . 17

Majority Student Performance 17
Minority Student Performance 24

IV. COMPARISONS OF MAJORITY AND MINORITY STUDENTS 31

Visual Comparisons 31
Adjustment of Minority and Majority Predictions for Comparability 38
A Single Equation for Majority and Minority Students 47

V. CONCLUSIONS 50

Prediction Accuracy 51
Unexplainable Differences in Performance 52
Implications for the Medical School Admissions Process 54
Broader Implications 55

Appendix

A. DATA AND VARIABLES 58

B. EXPLORATORY DATA ANALYSIS 64

School-by-School Analysis for Majority Students . . . 64
Combining Majority Students Across Schools 67
Minority Students 68
A Single Equation for Majority and Minority Students 69

C. PREDICTION ERRORS FOR COMPARING MINORITY AND MAJORITY STUDENT PERFORMANCE 72

LIST OF TABLES

1. Minority Student Enrollment in U.S. Medical Schools 2

2. Minority Student Retention 4

3. Students Repeating the Academic Year 5

4. Results of Regression Analysis of Majority Student Performance on National Board Exams 18

5. Results of Regression Analysis of Minority Student Performance on National Board Exams 25

6. Distribution of Minority and Majority Student Characteristics for Part I Exam Takers 41

7. Comparable Minority and Majority Student Predictions and Associated Errors for Mean Part I 43

8. Distribution of Minority and Majority Student Characteristics for Part II Exam Takers 45

9. Comparable Minority and Majority Student Predictions and Associated Errors for Mean Part II 46

10. Results of Regression Analysis for Combined Majority and Minority Student Performance on National Board Exams . 48

A-1. Majority Student Characteristics Means and Standard Deviations 62

A-2. Minority Student Characteristics Means and Standard Deviations 63

B-1. Admission Selectivity Index 66

LIST OF FIGURES

1. Mean Part I Scores: Predicted vs. Actual 32

2. Mean Part II Scores: Predicted vs. Actual 33

3. Biochemistry Scores: Predicted vs. Actual 34

4. Surgery Scores: Predicted vs. Actual 35

5. Medicine Scores: Predicted vs. Actual 36

6. Psychiatry Scores: Predicted vs. Actual 37

7. Mean Part I Predictions 40

8. Mean Part II Predictions 40

ACKNOWLEDGMENTS

This research would not have been possible without the trust and active cooperation of the deans of the medical schools in our sample. We could not have obtained and interpreted the data from student records without the assistance of the student affairs deans and administrative personnel responsible for student records.

Wendy Cooper coordinated the overall data collection effort for this study and abstracted literature relevant to this study. A review of the latter is contained in a separate report. Dan Relles provided invaluable advice on the problems of statistical computing. Without the benefit of the STATLIB software, of which Relles is co-developer, we might still be doing the computations for this report.

Joseph M. Brown, project officer for this contract, raised important issues throughout the course of this study and provided helpful criticism. Stephen Carroll and Finis Welch, Rand colleagues, Edward Ignall and William Schwartz, Rand consultants, and L. F. Krystynak of the Health Resources Administration provided valuable criticism of an earlier draft. Helen Turin edited the final report.

We are, of course, responsible for any errors of analysis and interpretation that remain.

I. INTRODUCTION

This report deals with the prediction of performance of minority and non-minority students in medical school to aid admissions committees in evaluating medical school applicants with very different backgrounds. Another purpose is to assist third parties in assessing the effectiveness and efficiency of affirmative action programs.

Before the advent of affirmative action programs, predicting medical school performance was of only limited interest. Competition for medical school places was so strong that admissions committees had little difficulty choosing applicants whose academic credentials showed that they could satisfactorialy complete the course of study. Few minority applicants had the academic credentials to be admitted, and those who did performed in a manner indistinguishable from other students.[1] Admission to medical school almost assured the student--minority or majority--an M.D. degree. Of course, not all students performed equally well, but little of the difference in performance could be attributed to measurable preadmission characteristics available to admissions committees.

At the end of the 1960s, U.S. medical schools faced new problems in predicting student performance as they began affirmative action programs targeted at certain minority groups--principally blacks, American Indians, Mexican Americans, and mainland Puerto Ricans. By the mid-1970s, minority recruitment and admissions had become a major concern of medical schools throughout the country. The results have been impressive as measured by the growth in minority medical student enrollment (see Table 1). There is intense competition for the academically highly qualified minority applicant, and many school resources are devoted to screening a much larger group of less highly qualified minority applicants to find those who are able and motivated to complete the demanding medical school curriculum. The schools have admitted minority applicants who are academically less well prepared to deal with the curriculum than most majority applicants.[2] Therefore it is hardly surprising that there is higher attrition among minority

Table 1

MINORITY STUDENT ENROLLMENT IN U.S. MEDICAL SCHOOLS

	Total Enrollment		First Year Enrollment		Graduates	
	No.	% of Total	No.	% of Total	No.	% of Total
1969-1970						
Black	1038	2.8	449[b]	4.3[b]	165	2.0
American Indian	16	a	5[b]	0.1[b]	4	.1
Mexican-American	92	0.2	45[b]	0.4[b]	10	.1
Puerto-Rican	34	0.1	11[b]	0.1[b]	9	.1
1970-1971						
Black	1525	3.8	708[b]	6.2	180	2.0
American Indian	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
Mexican American	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
Puerto-Rican	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
1971-1972						
Black	2090	4.0	890	7.2	229	2.4
American Indian	n.a.	0.6	21	n.a.	n.a.	n.a.
Mexican American	n.a.	n.a.	117	n.a.	n.a.	n.a.
Puerto-Rican	n.a.	n.a.	33	n.a.	n.a.	n.a.
1972-1973						
Black	2593	5.5	858	6.4	341	3.3
American Indian	68	0.1	33	0.2	8	.1
Mexican-American	380	0.8	151	1.1	39	.4
Puerto-Rican	98	0.2	44	0.3	10	.1
1973-1974						
Black	2961	5.9	870	6.4	511	4.5
American Indian	93	0.2	33	0.2	3	a
Mexican-American	512	1.0	170	1.2	79	0.7
Puerto-Rican	140	0.3	48	0.3	19	0.2
1974-1975						
Black	3396	6.3	950	6.5	638	5.0
American Indian	161	0.3	70	0.5	22	0.2
Mexican-American	638	1.2	196	1.4	110	0.9
Puerto-Rican	168	0.3	63	0.4	28	0.2

Source: *Journal of the American Medical Association.*

[a]Less than 0.05%.

[b]Includes students repeating the first year. All other data on first year students excludes repeaters.

students, and more repeat part of the curriculum than is the case for majority students (see Tables 2 and 3).

The central concern of admissions committees is how applicants will perform as physicians. The extensive literature on the prediction of medical school performance has been reviewed by Cuca, Sakakeeny and Johnson.[3] Most of the research has focused on the use of preadmission characteristics to predict which applicants to medical school will become good students and practitioners of medicine. The relevance of medical education's performance measures to medical practice has been a source of ongoing controversy.[4] Some researchers have questioned the validity of both the predictor variables (e.g., Medical College Admission Tests, college grade point average) and performance measures (e.g., medical school grades, National Board scores) because of the weak statistical relationship between them.[5] Because almost all studies have used data from a single medical school or no more than several schools, sample size has limited the application of multivariate statistical methods to these problems.

Already uneasy about their criteria for selecting future physicians from a fairly homogeneous white applicant pool, medical school admissions committees in recent years have had to recruit and select students from minority groups with different cultural and educational backgrounds. Most of the limited literature on affirmative action in medical school either describes the results of one or more experimental admissions programs or assesses affirmative action progress on the basis of aggregate data.[6] The unresolved issues of performance criteria and prediction accuracy have become intertwined with issues of social value and legal rights.[7]

In the recent Bakke decision, the U.S. Supreme Court held that medical schools may take racial factors into account in admissions decisions.[8] However, the Court also held that a particular admissions process employed by the University of California at Davis violated the Equal Protection Clause of the Constitution because it used rigid minority quotas effectively to exclude white applicants from consideration, irrespective of their other qualifications.

Table 2

MINORITY STUDENT RETENTION

	Black	American Indian	Mexican American	Puerto Rican	All Students
Admitted 1971-1972	758	21	117	33	10,962
Number retained June 74	649	21	110	30	10,500
Percent retained	86	100	94	91	96
Admitted 1972-1973	838	30	140	37	12,045
Number retained June 75	729	27	134	35	11,751
Percent retained	87	90	96	95	98
Admitted 1973-1974	908	37	167	48	12,393
Number retained June 75	790	31	157	47	12,066
Percent retained	87	84	94	98	97
Admitted 1974-1975	934	63	203	60	12,892
Number retained June 75	886	62	198	59	12,750
Percent retained	95	98	97	98	99

Source: *The Journal of the American Medical Association*, December 29, 1975, p. 1339.

Table 3

STUDENTS REPEATING THE ACADEMIC YEAR

	Black	American Indian	Mexican American	Puerto Rican (Mainland)	All Other Students
1973-1974					
First year class					
Enrolled	995	41	178	56	12,642
Repeating	120	9	8	8	142
Percent repeating	12.1	19.5	9.5	14.2	1.1
All other classes					
Enrolled	1966	52	334	126	33,477
Repeating	102	3	18	5	202
Percent repeating	5.2	5.8	5.4	4.0	0.6
1974-1975					
First year class					
Enrolled	1117	72	231	71	13,472
Repeating	161	2	32	7	162
Percent repeating	14.4	2.8	13.9	9.9	1.2
All other classes					
Enrolled	2279	89	407	97	36,239
Repeating	136	7	33	7	229
Percent repeating	7.2	8.1	8.1	7.2	0.6

Source: *The Journal of the American Medical Association*, December 29, 1975.

The problem for medical schools then is to administer affirmative action admissions programs that are fair and equitable to *individual* minority *and* majority applicants. Fairness and equity will inevitably be matters of judgment. However, admissions committees will surely--and the Bakke decision suggests they must--heed information about individual applicants that appears relevant to their eventual performance as a physician.[9]

This report is concerned with medical student scores on two standardized tests, Parts I and II of the National Board of Medical Examiners (NBME) exam.[10] In particular, we are concerned with predicting performance on these two exams for minority and non-minority students and with determining what factors, if any, account for differences between the two groups.

The NBME exam provides a measure of student performance--or more appropriately, educational achievement. Two parts of this three part standardized test are administered to medical students according to the policies of each individual school.[11] Successful completion of the first two parts of the exam during medical school and the third part during the internship (or the first year of graduate medical education) is accepted for licensure by all but three states.[12] Parts I and II are widely used by schools to assess student achievement, and in some schools passing is a prerequisite to promotion. Although the same test is given to students in all schools and scores are reported in standardized form, the meaning of a particular score varies across schools because the test is taken at different points in the medical school experience at different schools and because the schools use the exams in different ways to evaluate student programs.[13]

We do not make any judgments about the validity of the NBME exam for predicting how well students will perform as practicing physicians.[14] Such judgments are implicit in the importance that medical schools attach to the exams, and that importance varies substantially across schools. Notwithstanding this variation, the widespread use of the NBME exam makes it the most satisfactory and easily interpretable basis for comparisons of medical student performance involving multiple schools.

Section II describes the data and the statistical methodology used in our analysis. In Section III, we present the results of our analysis of factors that predict performance on the National Board exams using separate models for majority and minority students. The structures of the models are similar, but the parameters are different. In Section IV, we examine the implications of using a single model to predict *both* minority and majority student performance and discuss several methods for comparing both groups. A final section provides a nontechnical interpretation of the results and discusses some implications of our findings, particularly as they relate to medical school admissions policies.[15]

Footnotes--Section I

[1]The majority of blacks were enrolled in Howard and Meharry, two predominantly black medical schools. In 1968-69, 63 percent of the black medical students were enrolled in these two schools. By 1970-71, the increase in other schools' enrollment of blacks had caused this proportion to drop to 33 percent.

[2]Tables A-1 and A-2 give means and standard deviations for minority and majority students in the classes of 1975 and 1976 from the nine medical schools that are the subject of analysis in this report.

[3]J. M. Cuca, L. A. Sakakeeny, D. G. Johnson, *The Medical School Admissions Process: A Review of the Literature, 1955-1976,* Washington, D.C., Association of American Medical Colleges, 1976.

[4]A. R. Barro, "Survey and Evaluation of Approaches to Physician Performance Measurement," *Journal of Medical Education Supplement,* Vol. 48, November 1973, pp. 1051-1093; *Evaluation in the Continuum of Medical Education,* National Board of Medical Examiners, Philadelphia, June 1973.

[5]H. G. Gough, W. B. Hall, "The Prediction of Academic and Clinical Performance in Medical School," *Research in Higher Education,* Vol. 3, 1975, pp. 301-314; R. E. Roemer, "Nine Year Validity Study of Predictions of Medical School Success," *Journal of Education Research,* Vol. 59, No. 4, December 1965, pp. 183-185; M. A. Fredericks, P. Mundy, "The Relationship Between Social Class, Average Grade in Colleges, Medical College Admission Test Scores, and Academic Achievement of Students in Medical School," *Journal of Medical Education,* Vol. 42, February 1967, pp. 126-133; H. G. Gough, W. B. Hall, R. E. Harris, "Admissions Procedures as Forecasters of Performance in Medical Training," *Journal of Medical Education,* Vol. 38, December 1963; P. B. Price, C. W. Taylor, D. E. Nelson, *Measurement and Predictors of Physicians Performance: Two Decades of Intermittently Sustained Research,* Salt Lake City, LLR Press, 1971.

[6]D. A. Evans, et al., "Traditional Criteria as Predictors of Minority Success in Medical School," *Journal of Medical Education*, Vol. 50, October 1975, pp. 934-939; H. J. Simon, J. W. Covell, "Performance of Medical Students Admitted Via Regular and Admission Variance Routes," *Journal of Medical Education*, Vol. 50, March 1975, pp. 237-241; J. C. Plagge, et al., "Increasing the Number of Minority Enrollees and Graduates: Medical Opportunities Program," *Journal of Medical Education*, Vol. 49, August 1974, pp. 736-745; D. G. Johnson, V. C. Smith, Jr., S. L. Tarnoff, "Recruitment and Progress of Minority Medical School Entrants, 1970-1975," *Journal of Medical Education Supplement*, Vol. 50, July 1975, pp. 713-755; B. C. Sleeth, R. I. Mishell, "Black Under-Representation in United States Medical Schools," *New England Journal of Medicine*, Vol. 297, 1976, pp. 1146-1148.

[7]B. Davis, "Academic Standards in Medical Schools," *New England Journal of Medicine*, Vol. 294, 1976, p. 1118; Y. B. Burke, "Minority Admissions to Medical Schools: Problems and Opportunities," *Journal of Medical Education*, Vol. 52, September 1977, pp. 731-734; K. Abarbanel, "Is a Relapse Ahead for Minority Medical Education?" *Foundation News*, November/December 1976, pp. 11-28; F. J. Ramsely, "The Medical School Admissions Dilemma: Damned If You Do, Damned If You Don't," *Journal of American Medical Association*, Vol. 237, pp. 1093-1094.

[8]University of California Regents v. Bakke, No. 76-811.

[9]For further discussion of minority and majority admissions, see G. M. Carter, D. S. C. Chu, J. E. Koehler, R. L. Slighton, A. P. Williams, *Federal Manpower Legislation and the Academic Health Centers: An Interim Report*, R-1464-HEW, The Rand Corporation, April 1974, pp. 21-31; A. P. Williams, W. D. Cooper, C. L. Lee, *Factors Affecting Medical School Admission Decisions for Minority and Non-Minority Applicants: A Comparative Study of Ten Schools*, R-2030-HEW, The Rand Corporation, 1978; and J. E. Rolph, A. P. Williams, C. L. Lee, *The Effect of State of Residence on Medical School Admissions*, R-2014-HEW, The Rand Corporation, 1978.

[10]For a discussion of the NBME exam and its present and future role, see *Evaluation in the Continuum of Medical Education* (Report of the Committee on Goals and Priorities of the National Board of Medical Examiners), Philadelphia, 1973.

[11]Until 1968, Part I--dealing with basic science courses--was administered at the end of the second year of study and Part II--dealing with clinical disciplines--was administered near the completion of the fourth year of medical school. Since 1969, students taking the exams for credit toward certification have been admitted to the exam during any year of medical school without prerequisite courses. In general, school policies determine when students will take the exam and what role it plays in the evaluation process.

[12]It is based on 1973 data. *Evaluation in the Continuum of Medical Education*, p. 33.

[13]Each medical school's use of the National Board exams is described in Association of American Medical Colleges, *Curriculum Directory*, various years, Washington, D.C.

[14]Recent analysis shows that the scores on the NBME Part I and II exams are among the best predictors of performance on the written portion of the American Board of Internal Medicine specialty certification exam. See Robert M. Bell, *Medical School and Physician Performance: Predicting Scores on the American Board of Internal Medicine Exam*, R-1723-HEW, The Rand Corporation, September 1977.

[15]A separate report briefly reviews the literature related to the minority student admissions process and affirmative action program assessment in medical schools. Cooper, Lee, and Williams (1978).

II. METHODOLOGY

THE DATA

The data for this study were drawn from the student records of ten medical schools selected to be representative of the broad spectrum of U.S. medical schools.[1] We chose the graduating classes of 1975 and 1976 so as to have the most recent data and obtain the largest possible number of minority students for our analysis. No two schools' student records are identical; fortunately they contain many common data elements. Although we coded a number of data elements from school student records that were not common to all schools, for purposes of the study we restricted our analysis to variables that were common to all school records or whose values could easily be estimated or calculated from other data in a student's record (e.g., undergraduate GPA from science GPA, nonscience GPA, and the units of each). The data on premedical school characteristics were the Medical College Admissions Test scores, undergraduate GPAs, undergraduate school attended, sex, and race. Data on medical school characteristics were class to which the student was admitted and dates of taking Part I and Part II of the National Board of Medical Examiners examination. Data on medical school performance come from the results of Part I and Part II of the NBME examination as well as whether the student repeated at least one year of medical school.

Almost all medical schools in the United States require applicants to take the MCAT, and the MCAT scores were available from the records of all ten schools in our sample. It is a standardized test consisting of four parts: Science, Quantitative Ability, Verbal Ability, and General Information. Scores for each of the four parts are reported separately in the common form for such tests.[2] That is, standardized scores range from 200 to 800 with a mean of 500 and a standard deviation of 100.

The National Board of Medical Examiners administers a certifying examination consisting of three parts. The first two are given during medical school; Part III is given during the first year of graduate medical education (internship).

Part I of the NBME exam measures a student's knowledge of material in the standard basic science courses of anatomy, behavioral science, biochemistry, microbiology, pathology, physiology, and pharmacology. The medical school course material for these subjects are usually a combination of lecture, small section, and laboratory sessions.[3] Thus, the basic science course material and the teaching approach are both familiar to medical students from their undergraduate education, and this first part of medical school may reasonably be viewed as an extension of premedical training.[4]

The Part II NBME exam, unlike Part I, covers material that few students have been exposed to before medical school. The six exam sections deal with basic problems of diagnosis and patient treatment encountered in the different medical specialties: internal medicine, obstetrics and gynecology, pediatrics, psychiatry, public health, and surgery. In general, very little of the material on which Part II is based is presented to the student in the normal lecture or laboratory format. Most is either presented at the bedside, in small group discussion of diagnosis and treatment of patients in the clinical service, and at clinical conferences; or the student learns it through independent study of clinical journals and texts. Both the material and the learning process are so different from what the student has encountered before medical school that the Part II exam cannot be viewed merely as another in a series of tests measuring the absorptive capacity for knowledge presented by conventional educational means.[5] Like the MCAT, the NBME examination results are reported in standardized scores.

Although the NBME examination is a standardized test instrument, its use varies across medical schools and among students. For the academic year 1975-1976, 29 of the U.S. medical schools (one from our sample) did not require students to take the exam; the remaining schools required the exam; 34 schools (two from our sample) require students to pass the National Board examination as a condition of promotion.[6] Schools give Part I as early as the first year of medical school and as late as the end of the second year; Part II is given as early as the second and as late as the fourth. Students who are

required or chose to take the examination may do so as candidates or as noncandidates for certification.

STATISTICAL MODELING

The simplest way to compare minority and majority student examination scores is to compare the average minority student score with the average majority student score for each of the sections of Part I and Part II of the National Boards. If these averages were approximately equal within the limits of statistical fluctuation, we could conclude that the performance of minority students was comparable to that of the majority students. However, it is unreasonable to expect minority students, whose preadmission characteristics are on average well below those of the average majority students, to have National Board scores comparable to majority students' scores.

We have developed regression equations that allow us to control for the effect of preadmission characteristics on performance. One set of equations yields predictions of majority student scores on the National Boards as a function of their characteristics on entering medical school; another set does the same for minority students. An important aspect of affirmative action programs is that admissions committees use different criteria for admitting minority students and majority students. Different equations for majority and minority students might result from a combination of a number of factors. For example, the admissions committee might be doing a good job in selecting minority students who will perform well in a medical school setting in spite of nominally low qualifications. Also, the minority students themselves might be learning more from medical school on average than their majority classmates. Such results would be evidence that affirmative action programs are working to admit students whose true potential is greater than their academic records suggest.

The list of premedical school characteristics given above take numerical values with the exception of race and undergraduate school attended. For the minority student equations, race was coded as black, Asian, and other (mostly Mexican-Americans and Puerto Ricans) using dummy variables. (A race variable was not needed for the

majority student equation.) The student's undergraduate college was coded into ten groups using the classification of *Barron's Profile of American Colleges and Universities.*[7] Using entrance examination and high school record data on entering students and the proportion of applicants accepted, Barron's classifies colleges into seven groups: most competitive, highly competitive, very competitive, competitive, less competitive, non-competitive, and special. Three categories (highly competitive, very competitive, and competitive) have "+" associated with some colleges to denote that these colleges are somewhat more selective than other schools in the same category. We made separate categories out of the "+" schools and included special schools with the unclassified schools. We used these classifications as proxy variables for the quality of the undergraduate school from which the medical students graduated. In addition, a dummy variable for being a member of the class of 1976 was created, as was a dummy variable to indicate whether a student had repeated a year of medical school. Finally, two variables give the length of time (in years) since medical school entry that a student took Part I and Part II of the NBME exams.

With the above candidates as explanatory variables, we did exploratory analysis aimed at getting minority student and majority student regression equations for six different dependent variables: mean Part I score, mean Part II score, score on the biochemistry section of Part I, and scores on the psychiatry, medicine, and surgery sections of Part II. Appendix A gives further details on the definitions of the above list of independent variables and dependent variables and the reasons for their selection.

We began by fitting regression equations to the scores for each of several sections of Part I and Part II of the NBME exam for majority students, separately for each of our nine medical schools.[8] We initially analyzed each school's data separately because of our impression that the educational processes were substantially different across schools but modified this view somewhat after some analysis. Although our final model included all the schools in a single equation, the analysis that led to our choice of independent variables

in the separate school equations carried over to the combined school equation.

After trying a number of transformations of the independent variables and interactions between them for predicting mean Part I scores, we concluded with quite a simple model. It is linear in the following independent variables: all four MCATs, the time lapsed between entry and taking the exam, undergraduate GPA, selectivity index,[9] dummy variables for repeating a year, being female, entering in 1971, and dummy variables to account for some missing values. A similar analysis carried out for Part II and for some of the subsections of both Part I and Part II reached the same conclusions.

Some similarities in the coefficients of the nine different school equations led us to pool all the students and attempt to find a single regression equation that would predict exam scores for majority students in all nine schools. We tried a number of specifications by choosing which variables were allowed separate school effects and which were not. For Part I, we found that *only one variable was needed to capture the separate school effects and the best one (highest R^2) was the school dummy variable.*[10]

Section III reports the results of the Part I and Part II regressions as well as the subsection regressions--biochemistry, medicine, surgery, psychiatry--for the equations using the school dummies and all other variables being the same across the nine schools.

There are 268 minority students who took Part I and 199 minority students who took Part II in our sample. The comparable sample sizes of majority students was 1,899 and 1,309. This small number of minority students means that we could explore only simple structures for predicting minority student performance. As a result, the *structure* (but not the coefficients) of our final minority score prediction equation relies as much on the analogous majority student exploratory data analysis as it does on the exploratory data analysis of the minority student scores. In our final minority regression equations reported in Sec. III, we used the same functional form as for the majority students.

A variable by variable comparison of the coefficients of the comparable regression equations for majority and minority students reveals that most of the estimated coefficients are consistent with the true coefficients being equal--i.e., most variables have *the same effect on the predicted scores of minority and of majority students.* This is as expected. Since majority and minority students are in the same schools, are taking the same courses, etc., differences in their regression equations should be only for variables that affect minority and majority students differently. In Sec. IV we give a single equation model for predicting both minority and majority scores as an alternative to a separate equation for each group. Our exploratory analysis for developing this single regression equation is given in App. B.

We discuss the fitted equations for both majority and minority students in Sec. III. In Sec. IV we use our regression equations to characterize the relative performance of minority and majority students on the NBME exams. Since these equations cannot predict exam performance perfectly, statistical formulas are used to calculate the probable error in our estimated scores when we compare minority and majority students. Appendix C describes the formulas used for these prediction errors.

Footnotes--Section II

[1]For a description of the selection process, see G. Carter, D. Chu, J. Koehler, A. Williams, *The Federal Government and Academic Medicine: The Effects of Federal Programs on Activities and Outputs*, R-1814-HEW, The Rand Corporation, 1978 forthcoming.

[2] The MCAT is an objective measure of general and specific academic aptitude which provides scores in four areas. The *Verbal Ability* subtest measures knowledge of vocabulary and ability to perceive verbal relationships. The *Quantitative Ability* subtest requires the application of basic mathematical principles. These two scores are directly related to an individual's general academic aptitude. The *Science* subtest samples the candidate's store of scientific information and of principles found to be important in preparing for the study of medicine. It is oriented toward predicting success in preclinical studies. Specifically, it is an indication of the capability of the examinee in areas usually covered by introductory courses in biology, organic and inorganic chemistry,

and physics. The fourth subtest, *General Information*, is a measure of overall cultural knowledge, including the social and behavioral sciences.

Association of American Medical Colleges, *Medical School Admission Requirements, 1977-78, United States and Canada*, Washington, D.C.

[3]There is a tendency toward reduction of the laboratory component in the basic science courses, and a few schools have computer assisted instruction for these courses.

[4]Robert Ebert has recently discussed this point and has suggested that the university maintain responsibility for medical science training for two years beyond completion of undergraduate training and thereafter pass responsibility of actual medical training (currently the last two years of medical school plus residency) to teaching hospitals. See Robert H. Ebert, M.D., "Medical Education in the United States," *Daedalus*, Winter 1977, pp. 171-84.

[5]We draw the distinction to differentiate the Part II test from a range of tests that students encounter, beginning with elementary school achievement tests and extending through SATs, MCATs, and Part I of the National Board exam. We make no pretense at having investigated the relationship of Part II performance to actual clinical competence in the component medical specialty parts.

[6]This information is summarized in *1975-1976 AAMC Curriculum Directory*, Association of American Medical Colleges, Washington, D.C., 1975.

[7]*Barron's Profiles of American Colleges*, Ninth Edition, Barron's Educational Series, Inc., Woodbury, New York, 1974.

[8]One of the ten schools in the study did not have any minority students who took the NBME exams and so was not included in the analysis.

[9]This selectivity index is a scale based on Rolph, Williams, and Lee (1977). Its values reflect the relative effect being from a college with a given Barron's index has on the probability of being admitted to medical school. See App. B for details.

[10]Additional variables did not reduce the unexplained variance by a statistically significant amount.

III. ANALYSIS OF MAJORITY AND MINORITY STUDENT PERFORMANCES

Although this report is more concerned with predicting performance of minority students in the medical school curriculum than with that of majority students, we first present the results of our analysis of majority student performance on both parts of the NBME exam, then our analysis of minority student performance. There are two reasons for this order. The larger size of the majority student population made it the more suitable for the exploratory data analysis and helped us to determine an appropriate structure for all our predictive models. Further, because there are many more majority than minority students, the point of reference that medical school administrators and faculty use in assessing minority student performance is their experience with majority students.

We present six separate prediction equations for minority and majority performance. The two equations for Part I predict the mean on all seven sections and the score on the biochemistry section. The four equations for Part II predict the mean of all six sections and the separate scores on the medicine, surgery, and psychiatry sections. The results of the regression analysis of majority student performance are presented in Table 4. The corresponding results for minority students are presented in Table 5.

MAJORITY STUDENT PERFORMANCE

Part I Mean

Because the first part of medical school may reasonably be viewed as an extension of premedical training, we would expect to be able to predict with fair accuracy the performance on the Part I exam using a student's preadmission characteristics. However, in absolute terms, our ability to predict mean Part I scores for majority students is limited. Although our multiple regression equation is highly statistically significant, we are able to explain only 31 percent of the variance. See first column of Table 4. It is not surprising that the science MCAT and undergraduate GPA are the strongest predictive

Table 4

RESULTS OF REGRESSION ANALYSIS OF MAJORITY STUDENT PERFORMANCE
ON NATIONAL BOARD EXAMS
(Standard deviations in parentheses)

	Part I		Part II			
Variable	Mean	Bio-Chemistry	Mean	Medical	Surgery	Psychiatry
Constant	30.221	170.750[c]	-75.282[b]	49.985	38.129	-17.561
	(27.692)	(31.261)	(35.964)	(38.520)	(41.405)	(37.365)
School 0	9.73[illegible]	18.148[a]	15.850	10.790	21.333	10.221
	(8.299)	(9.368)	(12.018)	(12.873)	(13.837)	(12.487)
School 1	-40.725[c]	-.718	-35.427[c]	-18.256[a]	-19.253[a]	-17.152
	(9.732)	(10.986)	(10.134)	(10.854)	(11.667)	(10.529)
School 2	41.316[c]	43.639[b]	21.671[b]	46.827[c]	54.379[c]	-14.124
	(13.161)	(14.857)	(9.532)	(10.209)	(10.974)	(9.903)
School 3	.795	-10.602	19.242[b]	49.683[c]	42.809[c]	16.190
	(10.319)	(11.649)	(12.262)	(13.134)	(14.117)	(12.740)
School 4	-10.792	9.923	5.943	11.228	11.734	-3.145
	(9.378)	(10.586)	(8.856)	(9.485)	(10.195)	(9.201)
School 6	-3.432	57.046[c]	-36.999[c]	20.753[b]	2.540	-25.652[b]
	(8.864)	(10.006)	(9.470)	(10.143)	(10.903)	(9.839)
School 7	-13.911	1.756	-33.151[c]	-37.273[c]	-10.517	-10.688
	(9.584)	(10.819)	(10.606)	(11.360)	(12.211)	(11.019)
School 8	8.752	60.062[c]	-22.8[illegible]	-25.582[b]	-20.771[a]	2.300
	(8.711)	(9.834)	(9.7[illegible]	(10.442)	(11.224)	(10.129)
Exam Date	-9.998[b]	-47.700[c]	20.57[illegible]	22.816[c]	25.821[c]	18.782[a]
	(4.453)	(5.027)	(7.550)	(8.087)	(8.692)	(7.844)
Missing Exam Date	1.860	-46.241[c]	55.807[b]	74.716[c]	81.848[c]	36.371
	(12.174)	(13.742)	(26.642)	(28.536)	(30.673)	(27.680)
Verbal MCAT	.094[c]	.035	.120[c]	.018	.067[a]	.257[c]
	(.028)	(.032)	(.035)	(.038)	(.041)	(.037)
General Information MCAT	.052[a]	.006	.157[c]	.087[b]	.074[a]	.196[c]
	(.030)	(.034)	(.038)	(.040)	(.043)	(.039)
Quantitative MCAT	.032	.079[c]	.074[b]	.029	.000	.039
	(.027)	(.030)	(.033)	(.036)	(.039)	(.035)
Science MCAT	.436[c]	.363[c]	.352[c]	.325[c]	.318[c]	.188[c]
	(.029)	(.032)	(.036)	(.038)	(.041)	(.037)
Undergraduate GPA	35.409[c]	28.280[c]	31.169[c]	26.741[c]	31.303[c]	24.428[c]
	(4.646)	(5.245)	(5.400)	(5.784)	(6.217)	(5.610)
Missing GPA	117.830[c]	80.902[c]	112.010[c]	95.228[c]	101.210[c]	86.207[c]
	(17.365)	(19.603)	(19.839)	(21.249)	(22.840)	(20.612)
Year 1975	12.226[c]	6.973	-8.520[a]	-9.211[a]	-5.542	-23.730[c]
	(4.105)	(4.634)	(5.190)	(5.558)	(5.975)	(5.392)
Repeat	-2.573	4.176	12.526	-3.772	10.196	24.241
	(10.646)	(12.018)	(14.708)	(15.753)	(16.933)	(15.281)
Female	-17.433[c]	-21.594[c]	1.845	-12.589[b]	-21.457[c]	20.841[c]
	(4.794)	(5.412)	(6.003)	(6.430)	(6.912)	(6.237)
Selectivity Index	28.774[c]	29.373[c]	21.700[b]	10.355	5.244	26.827[c]
	(7.604)	(8.583)	(8.906)	(9.540)	(10.254)	(9.253)
Estimated Standard Deviation of Regression	72.08	81.370	75.90	81.29	87.38	78.85
R^2	.313	.277	.317	.262	.229	.270
F	42.814[c]	35.936[c]	31.327[c]	23.865[c]	20.008[c]	24.521[c]
Degrees of Freedom	(20,1878)	(20,1878)	(20,1348)	(20,1348)	(20,1348)	(20,1348)
Sample Size	1899	1899	1369	1369	1369	1369

Notes: a. Significant at the .10 level.
b. Significant at the .05 level.
c. Significant at the .01 level.

variables in terms of statistical significance. For example, we see from the coefficient on science MCAT (.456) that a 100 point difference in that score, other things equal, results in an estimated 44 point difference in the total NBME score. Similarly, a 1.0 difference in GPA accounts for 35 points difference in the NBME score. By contrast, although verbal MCAT is related to NBME Part I scores in a statistically significant manner, the model indicates that a 100 point difference in the verbal MCAT would account for only a 9 point difference in the mean Part I score. Students with missing GPAs perform slightly less well than the average students, as though their GPA were 3.33 (117.8 ÷ 35.4) rather than the mean GPA of 3.4.

The quality of the undergraduate school as measured by the selectivity index is positively associated with the Part I score, and the relationship is statistically significant. However, the maximum effect of undergraduate school quality (after controlling for other factors) covers a span of only 29 points between the most and least selective schools. This undergraduate school effect is dwarfed by some of the *medical school* effects on NBME score. When we control for other characteristics, school 1 students score on average an estimated 41 points below school 9 students (the default value), and school 2 students score on average 41 points above school 9 students.

Female students score on average an estimated 17 points below male students with equivalent backgrounds; this difference is statistically significant. Other categorical variables with statistically significant coefficients show that, other things being equal, students taking the exam a year earlier in the curriculum score about 10 points higher, and a member of the class of 1975 scores 12 points higher on the exam than his 1976 counterpart.

The equation for majority mean Part I scores offers no real surprises. The variables with the greatest predictive power are the ones expected, and in no case is the sign of a coefficient the opposite of what our limited theory would suggest.

Part I Biochemistry

Since the biochemistry section score is a component part of the mean score on the Part I exam, it is not surprising that the two are generally similar. The predictive power of the model is somewhat less for biochemistry (R^2 = .28) than for the Part I mean.

The most interesting difference is in the "school effects" between the two exams. An examination of the school dummy variable coefficients shows that the effect for school 2 (as compared to school 9) is almost identical for mean Part I and Part I biochemistry. However, there are strong differences among other schools. For example, school 1 students with comparable preadmission characteristics scored 41 points below school 9 students on mean Part I, but their scores are indistinguishable on the biochemistry section.

Although it is beyond the scope of this study to interpret the cause of school-to-school differences, some generalizations can be made without further analysis. It is obvious that there are significant differences (both statistical and substantive) in performance among students at the nine schools after we have controlled for their preadmission characteristics, when they took the exam, and so on. However, the differences between mean Part I and Part I biochemistry suggest that what we have called a "school effect" might be better termed a "curriculum effect." That is, students at one school may score high on one part of the exam and low on another. These differences may be attributed to the emphasis, quality of instruction, timing in basic science course sequence, or other factors associated with the subject area. Whatever the reasons for the differences, it is incorrect to attribute school-to-school differences in performance on the NBME exam merely to the importance each school attaches to the exam as a whole.

Science MCAT and undergraduate GPA are the two best predictor variables for the biochemistry section, as was the case for mean Part I. The quantitative MCAT is statistically related to the biochemistry score ($p < .01$), and the verbal MCAT is not. The opposite was true for mean Part I. Other things equal, the later students take the exam, the less well they perform. A year's delay is associated with an

estimated loss of 48 points on the biochemistry section, perhaps because most of the biochemistry instruction is given during the first year of the basic science curriculum at the medical schools in our sample. The large negative coefficient for missing exam date suggests that many of these missing dates may be late ones. The effects of the other categorical variables on performance do not seem to differ in important ways from those observed for mean Part I.

Part II Mean

An examination of the prediction model for Part II mean reveals two general characteristics. First, the predictive power is almost identical in terms of total explained variance to that of Part I mean. Second, considerably more variables contribute statistically significant predictive power to the Part II model than the Part I [illegible]

Although science MCAT and undergraduate GPA are the strongest predictor variables, all MCAT parts are statistically significant predictors of mean Part II performance. The general information MCAT is second only to science among the MCAT parts; the former was not significantly related to mean Part I or Part I biochemistry performance.[1] A 100 point difference in the general information score is roughly equivalent in predicting mean Part II to a .5 difference in GPA.

Examination of the school dummy variables shows that scores by students from five schools (1, 2, 6, 7, 8) differ in a statistically significant way from those of comparable students in school 9 (the default value). The strength and pattern of these efforts are different from those observed on the Part I tests.

The later the exam is taken, the better, on average, students perform. This is contrary to what we observed for the Part I exam, but it is not surprising. The material covered on Part II is generally presented during full time "rotations" on a particular clinical service, such as medicine or surgery. The later in the training the Part II exam is taken, the more rotations students will have had and the more familiar they will have become with the problems of individual specialties and of clinical medicine in general. Those with missing exam dates score approximately the the same on average as other students.

Examining the other categorical variables, we see that the effect of sex observed in Part I disappears, as does the difference between the classes of 1975 and 1976. The selectivity of the undergraduate school is also less important than it was for the Part I tests.

Part II Medicine

As has been the case with the three equations described previously, science MCAT and undergraduate GPA are the two strongest predictor variables for the medicine section of the Part II exam. General information is the only other MCAT with a statistically significant coefficient, but its effect is not large--an estimated 9 points difference on the medicine score from a 100 point difference on the general information score.

A comparison of school dummy variables with those for mean Part II shows some significant differences supporting the hypothesis that "school effects" are more properly viewed as curriculum effects. That is, they differ from subject area to subject area.

The effect of undergraduate school quality is no longer statistically significant, although the sign remains positive. However, women score an estimated 13 points below comparable men, but this difference is small and is barely significant at the .05 level.

Part II Surgery

The surgery equation for majority students has less predictive power than any of the other equations in this report. Science MCAT and GPA are the two most statistically significant predictor variables. None of the other coefficients for variables that measure premedical school academic preparation is statistically significant, including selectivity of the undergraduate school attended.

Only two of the medical school dummy variables have significant predictive power, suggesting that school-to-school differences in surgery training may be less than in other subjects. We do observe a positive effect of taking the exam later, that is consistent across all parts of Part II. We also observe that women do significantly less well than men whose other characteristics are comparable. This

is consistent with analysis that shows women are less likely to chose specialty training in surgery than in other fields.[2]

Part II Psychiatry

The equation for predicting Part II psychiatry scores is different in a number of ways from any others. The overall predictive power of the model is roughly the same as in many of the other exam parts, but the relative predictive power of the variables has a different pattern from any of the other models.

Although their coefficients are statistically significant, science MCAT and GPA are not the dominant predictor variables as they were for other sections of the NBME exam. Verbal MCAT is a better predictor than either. The coefficient for general information MCAT is of roughly the same size and statistical significance as the science MCAT and is a better predictor than GPA.[3]

The effect of undergraduate school quality appears to be significantly stronger for psychiatry than for other components of the Part II exam. By contrast, the effects of medical school curriculum across schools is the least for any of the exam parts; only students from school 6 have scores that are significantly different from those of school 9 students.

Several differences in the effects of categorical variables on psychiatry scores are noteworthy. Women score statistically significantly higher than men; this is the only exam for which we have observed that effect.[4] Students from the class of 1976 perform significantly better than do members of the class of 1975. The coefficient on missing GPA suggests that students from undergraduate schools with pass/fail grading systems (those with missing GPA) perform at least as well as on the psychiatry subtest as those from schools with conventional grading systems.

MINORITY STUDENT PERFORMANCE

Here we describe the prediction models for minority student performance, concentrating on characteristics that distinguish minority student models from those that apply to majority students. The summary statistics for minority and majority students show not only that the minority means are lower than majority means but also that there is greater variability (the standard deviations are larger) in the former than the latter (see Tables A-1 and A-2). We may therefore expect predictive power of the minority equations to be somewhat better than that for the corresponding majority equations; that is, our equations have larger R^2 for the NBME scores of the minority sample than for the majority sample, if the effects of variables are generally comparable. This is the case for all exam parts. However, that does not necessarily mean we would expect to predict an individual minority student's score more accurately than an individual majority student's score. Such comparisons of prediction error for *individual scores* can most easily be made by examining the "estimated standard deviation of regression" entry in the tables for the corresponding majority and minority prediction equations.

Part I Mean

The model for minority mean Part I performance, shown in Table 5, clearly has considerable predictive power. As expected, the proportion of total variance explained is higher (R^2 = .58) than for the corresponding majority equation (R^2 = .31). We also observe that the estimated standard deviation for the minority equation is smaller than that for the majority equation. Thus, the minority equation yields more accurate predictions of scores for individuals as well as the group.

As was the case with the majority equation, undergraduate GPA and science MCAT have statistically significant coefficients. The notable difference between the two equations is the strength of quantitative MCAT in the minority equation. This variable, whose coefficient was not statistically significant in the majority equation, has stronger predictive power for minority performance than GPA and only slightly

Table 5

RESULTS OF REGRESSION ANALYSIS OF MINORITY STUDENT PERFORMANCE ON NATIONAL BOARD EXAMS
(Standard deviations in parentheses)

	Part I		Part II			
Variable	Mean	Bio-Chemistry	Mean	Medical	Surgery	Psychiatry
Constant	47.680	199.51[c]	-93.085	68.188	-36.468	38.798
	(58.226)	(58.121)	(75.423)	(86.756)	(91.317)	(87.575)
School 0	5.344	10.165	17.020	4.289	54.949	22.330
	(33.318)	(33.257)	(46.465)	(53.447)	(56.257)	(53.952)
School 1	-65.987[b]	-45.320	-51.651	-74.569[a]	-39.712	36.191
	(31.459)	(31.402)	(36.056)	(41.473)	(43.654)	(41.865)
School 2	54.279[a]	50.391	31.406	31.369	84.875[c]	10.349
	(31.773)	(31.716)	(25.416)	(29.235)	(30.772)	(29.512)
School 3	-13.487	-29.973	16.126	18.925	70.532[a]	60.364[a]
	(26.238)	(26.190)	(30.273)	(34.822)	(36.652)	(35.151)
School 4	-40.534	-1.253	-7.814	-16.240	15.916	20.384
	(29.379)	(29.316)	(30.133)	(34.661)	(36.483)	(34.988)
School 6	3.508	53.591[b]	-16.371	18.117	38.830	-10.420
	(26.082)	(26.035)	(26.510)	(30.493)	(32.096)	(30.781)
School 7	-9.177	-12.911	.914	-31.519	48.003	28.923
	(24.424)	(24.380)	(26.920)	(30.965)	(32.592)	(31.257)
School 8	27.468	65.113[b]	9.365	-17.098	14.199	73.976[b]
	(28.750)	(28.697)	(30.272)	(34.820)	(36.651)	(35.149)
Exam Date	-25.043[a]	-44.350[c]	14.813	19.162	22.073	-13.124
	(13.000)	(12.976)	(23.209)	(26.696)	(28.100)	(26.948)
Missing Exam Date	-54.980[a]	-81.462[c]	65.495	81.623	74.246	-62.601
	(29.516)	(29.463)	(78.470)	(90.261)	(95.007)	(91.114)
Verbal MCAT	.110	.155[b]	.261[c]	.355[c]	.243[b]	.347[c]
	(.076)	(.076)	(.091)	(.105)	(.111)	(.106)
General Information MCAT	.062	-.091	.085	-.164	.004	.193[a]
	(.078)	(.077)	(.095)	(.109)	(.115)	(.110)
Quantitative MCAT	.202[c]	.229[c]	.111[a]	.093	-.019	.048
	(.059)	(.059)	(.068)	(.078)	(.082)	(.079)
Science MCAT	.269[c]	.208[c]	.194[b]	.108	.241[b]	.127
	(.071)	(.071)	(.083)	(.096)	(.101)	(.097)
Undergraduate GPA	31.258[c]	20.592[a]	35.614[c]	28.956[a]	37.245[b]	12.710
	(11.421)	(11.400)	(13.495)	(15.522)	(16.338)	(15.669)
Missing GPA	92.293[b]	51.302	134.190[c]	118.210[b]	148.800[b]	49.958
	(41.144)	(41.069)	(49.442)	(56.872)	(59.862)	(57.409)
Year 1975	-3.561	-8.622	-4.770	-10.252	-14.656	-14.348
	(10.954)	(10.934)	(13.129)	(15.102)	(15.896)	(15.245)
Repeat	-20.429	-39.029[a]	-40.579	-10.367	-26.591	30.079
	(21.724)	(21.684)	(29.967)	(34.470)	(36.282)	(34.796)
Female	-20.185[a]	-23.294[b]	11.340	-11.033	-23.620	17.147
	(11.418)	(11.397)	(13.953)	(16.050)	(16.893)	(16.201)
Selectivity Index	-4.228	-2.825	9.180	14.437	1.160	-18.313
	(21.660)	(21.620)	(25.567)	(29.409)	(30.955)	(29.686)
Estimated Standard Deviation of Regression	68.70	68.57	68.79	79.12	83.28	79.87
R^2	0.575	0.560	0.537	0.402	0.428	0.421
F	16.591[c]	15.617[c]	10.316[c]	5.994[c]	6.653[c]	6.463[c]
Degrees of Freedom	20,245	20,245	20,178	20,178	20,178	20,178
Sample Size	266	266	199	199	199	199

Notes: a. Significant at the .10 level.
b. Significant at the .05 level.
c. Significant at the .01 level.

less than science MCAT. Since the strictly quantitative content of the Part I exam is small, this part of the MCAT may well be measuring characteristics associated with minority students' ability to learn during the very difficult circumstances of their first year or so in medical school. Perhaps this is abstract reasoning ability.

Another difference between the minority and majority equations is that undergraduate school quality (selectivity index) does not appear to affect minority performance when other factors are controlled for. This is somewhat surprising since it is a commonly held view that minority students have much to gain from attending the "better" undergraduate schools.[5]

The effects of other individual characteristics on the mean Part I score seem to follow the same general pattern for minority and majority students. These include the negative effect of being a woman, missing GPA, and waiting longer to take the test. The statistical significance of these effects is in every instance less than in the majority equation, and may be attributed to the much smaller minority sample size. There is no significant difference in minority performance between the classes of 1975 and 1976 as there was in the majority model.

The signs of the school dummy variable coefficients are the same for the minority and majority equations with two exceptions that are not significantly different from zero. The statistical significance of the minority coefficient is less, but this is because of sample size differences. However, the minority school coefficients are larger than the corresponding coefficients in the majority equation. Although these are not statistically significant, they do suggest that the relative strengths or deficiencies of a school's basic science program may have an amplified effect on the academically less well prepared minority student.

Part I Biochemistry

The equation for predicting minority student performance on the biochemistry exam has many characteristics in common with the mean Part I majority model. The proportion of predicted variance is higher and the estimated standard deviation smaller than for the corresponding majority equation.

The predictive power of the quantitative MCAT is the greatest of any variable in the minority equation, which is very different from the majority equation. The coefficient on the verbal MCAT score is also more statistically significant and contributes more predictive power to the equation than does GPA, suggesting that differences among minority students in assimilating the basic science material of the medical school curriculum may be due less to differences in their pre-medical school science training than to their ability to learn under the difficult circumstances of medical school.

As was the case with majority students, taking the test later generally implies lower scores. Those who repeat the exam (12 students in this case) performed worse than they would have been predicted to perform from only their other characteristics.

The sign of each of the coefficients for school dummy variables is the same as the corresponding variable for the mean Part I score. That was not the case for majority students. One explanation might be that "school effects" for minority students are greatly influenced by special tutoring programs to help them through basic science course work as a whole.

Part II Mean

The equation for predicting the mean score on the Part II exam for minority students is similar to the Part I minority equations in terms of overall predictive power. The proportion of the total variance explained is higher and the prediction standard deviation lower than for the corresponding majority model.

There are some modest differences in the effects of MCATs and GPA between the minority and majority models. Science MCAT has considerably less predictive power in the minority than in the majority equation, and the reverse is true for the verbal MCAT.

None of the coefficients for the categorical variables is statistically significant ($p < .10$) in the minority equation. Although most of the signs of the school dummy coefficients are the same in both minority and majority equations, the coefficients in the minority equation are too small to be of interest except for school 1 and

school 2. The levels of statistical significance in these two cases are so low as to make any discussion of school effects highly speculative.

Part II Medicine

In terms of predictive power, the relationship of the minority medicine equation to the majority medicine equation is similar to that in other parts of the NBME exam. However, we observe a different pattern in coefficient size and statistical significance between the two medicine equations for MCATs and GPA. The verbal MCAT, statistically insignificant in the majority equation, has the greatest predictive power in the minority equation. The general information MCAT score, which had a modest positive effect in the majority medicine equation, appears to have weak negative effect on minority performance on the medicine part of the exam. GPAs and the other MCATs have only weak positive effects in the minority equation.

The effects of school appear weak or uncertain for the minority equation. Although all but one of the school dummy coefficients have the same sign as in the majority equation, only the school 1 negative effect is strong and consistent enough to be even moderately statistically significant ($p < .10$).

Part II Surgery

The predictive power of the minority surgery equation is the least of any of the minority equations. However, its predictive power relative to the corresponding majority surgery equation is similar to other exam parts in terms of the differences in both R^2 and estimated standard deviations. The chief difference between the two equations is in the effects of the verbal and general information MCATs. The verbal MCAT has more predictive power for minority student performance. By contrast, the general information MCAT, which was a significant predictor of majority performance, is unrelated to minority performance.

The coefficients for the school dummy variables are generally larger and more statistically significant than those of the minority equations for mean Part II and medicine. In particular, minority

students from schools 2 and 3 performed much better on the surgery exam than would have been expected on the basis of their other characteristics. Since such "school effects" were absent or weak for other components of Part II, it seems reasonable to attribute the school effects to the effects of the surgery departments in the two schools. Both schools had large and statistically significant coefficients in the surgery equation for majority students, but the size of the minority equation coefficients was larger in both cases.

Part II Psychiatry

Unlike the equations for other components of the NBME exam, the minority equation for psychiatry is no better in predictive power than the corresponding majority equation. Although we are able to predict a larger proportion of the total variance with the minority psychiatry equation, the estimated standard deviation is slightly larger than in the case of the majority equation.

The verbal MCAT is clearly the best predictor variable for the minority psychiatry score, and the general information MCAT is the second best. Neither of the other two MCATs nor GPA is a statistically significant predictor.

Minority students from school 8 and school 3 perform significantly better than predicted on the basis of their other characteristics. Neither of these schools had a statistically significant coefficient in the majority equation, suggesting that the two psychiatry departments were uniquely successful with minority students. However, because there are few minority students in any particular school in the sample, one cannot draw strong inferences from the coefficients of these categorical variables.

Footnotes--Section III

[1]The MCAT test has been revised, and a new version was administered to the medical school class seeking admission in 1977. That version eliminates the general information part of the test.

[2]See A. P. Williams, G. M. Carter, D. S. C. Chu, S. B. Coleman, A. P. Massell, C. R. Neu, R. L. Rasmussen, W. H. Rogers, *The Effect of Federal Biomedical Research Programs on Academic Medical Centers*, R-1943-PBRP, The Rand Corporation, March 1976, pp. 30-37.

[3]A simple means for making this comparison is to multiply the standard deviations of each of the variables (from Table 6) by their coefficients. For the general information MCAT, one standard deviation difference accounts for an estimated 14 points on the psychiatry part of the exam (72 x .196). By contrast, one standard deviation difference in GPA accounts for an estimated 8 points difference in the psychiatry score (.34 x .244).

[4]It would not be surprising if women also scored better on the pediatrics or the obstetrics/gynecology parts of the Part II exam, since those have traditionally been popular specialties among women graduates. However, we did not include those parts of the exam in our analysis for reasons described in App. A.

[5]Since the selectivity index used here was derived from majority student data, an explanation of its lack of effect may be that the index is inappropriate for minority students.

IV. COMPARISONS OF MAJORITY AND MINORITY STUDENTS

In this section we present comparisons of predictions and actual scores of majority and minority students. We first show plots of predicted scores against actual scores to give a visual impression of how well the scores agree with the predictions. We then use statistical adjustment procedures to compare scores of majority and minority students with the same preadmission characteristics. Finally, we present an alternative model for predicting both majority and minority student scores with a single equation.

VISUAL COMPARISONS

Tables A-1 and A-2 show that average majority student scores are substantially higher than average minority student scores for mean Part I, mean Part II, and each of the subsections. Similarly, the means of the relevant preadmission characteristics (MCATs and GPAs) are substantially higher for majority students than for minority students. These averages give only a summary of the differences between the majority and minority distributions of scores. We first use plots to more fully display these differences. In Figs. 1 through 6 the predicted score is plotted on the horizontal axis and the actual score is plotted on the vertical axis for each student. All of minority student scores are plotted. Since there are too many majority students (1,899 in Part I and 1,369 in Part II) to plot without the points being so close together as to be indistinguishable, we selected a random sample of 200 majority students and plotted their scores.

Figure 1 plots predicted and actual mean *Part I* scores of majority and minority students. The points on the diagonal 45 degree line (the horizontal and vertical scales are different) correspond to students whose predicted scores are equal to their actual scores. Points above the diagonal line are underpredictions, and points below the diagonal are overpredictions. Minority students, indicated by dots, have lower average scores than majority students, indicated by circles, as shown by the tendency of the dots to be below the circles. Similarly, the

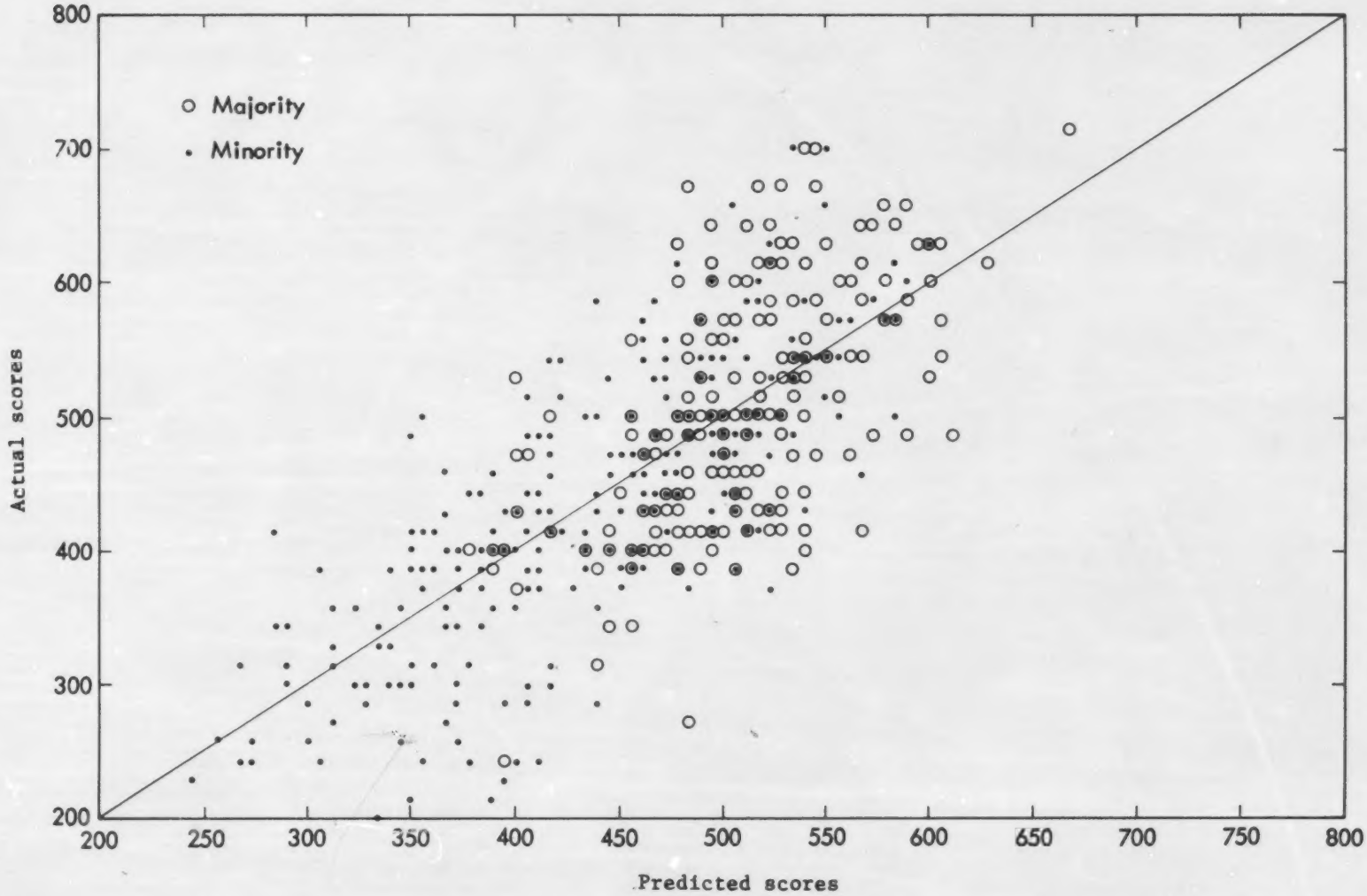

Fig. 1--Mean Part I scores: predicted vs. actual

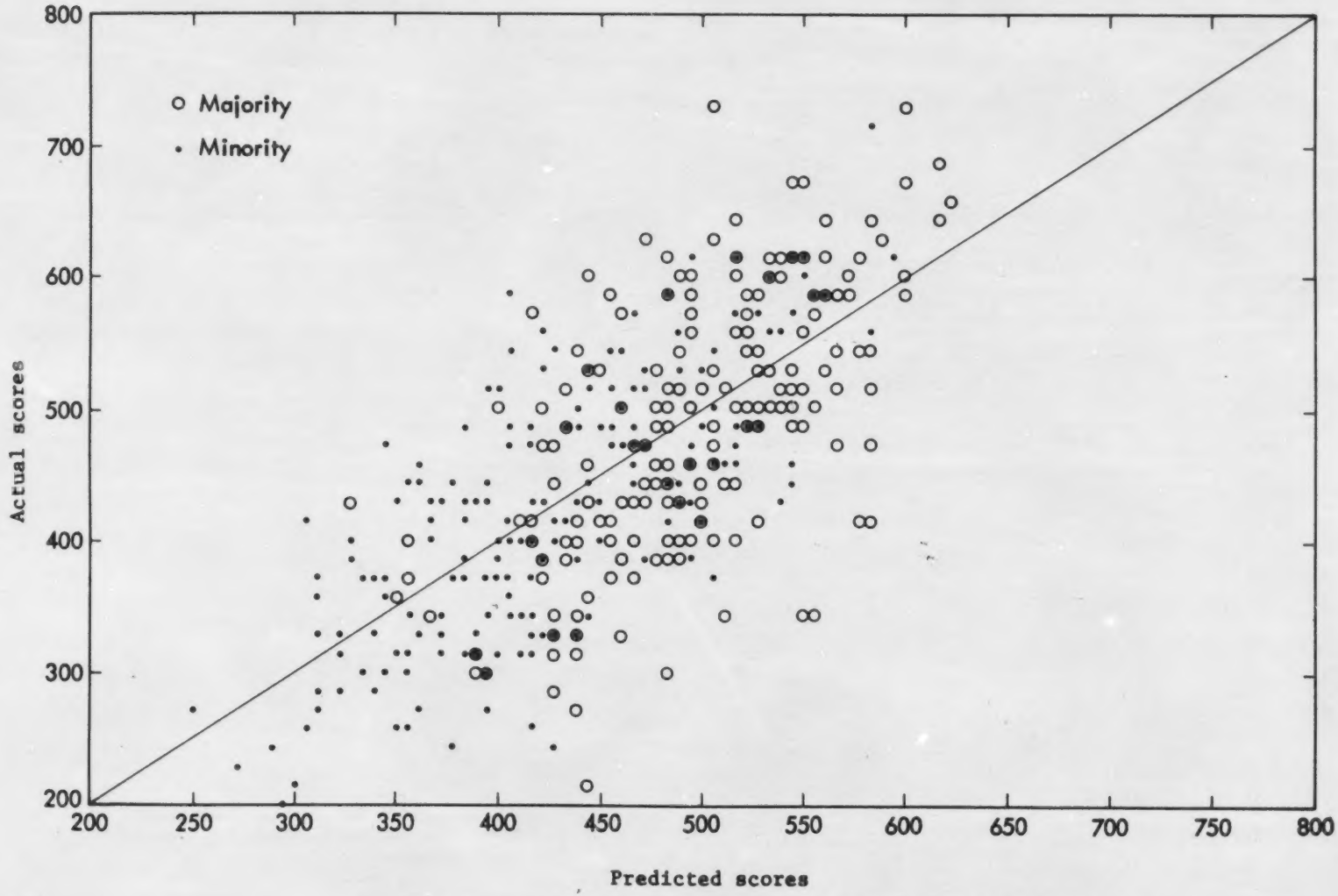

Fig. 2--Mean Part II scores: predicted vs. actual

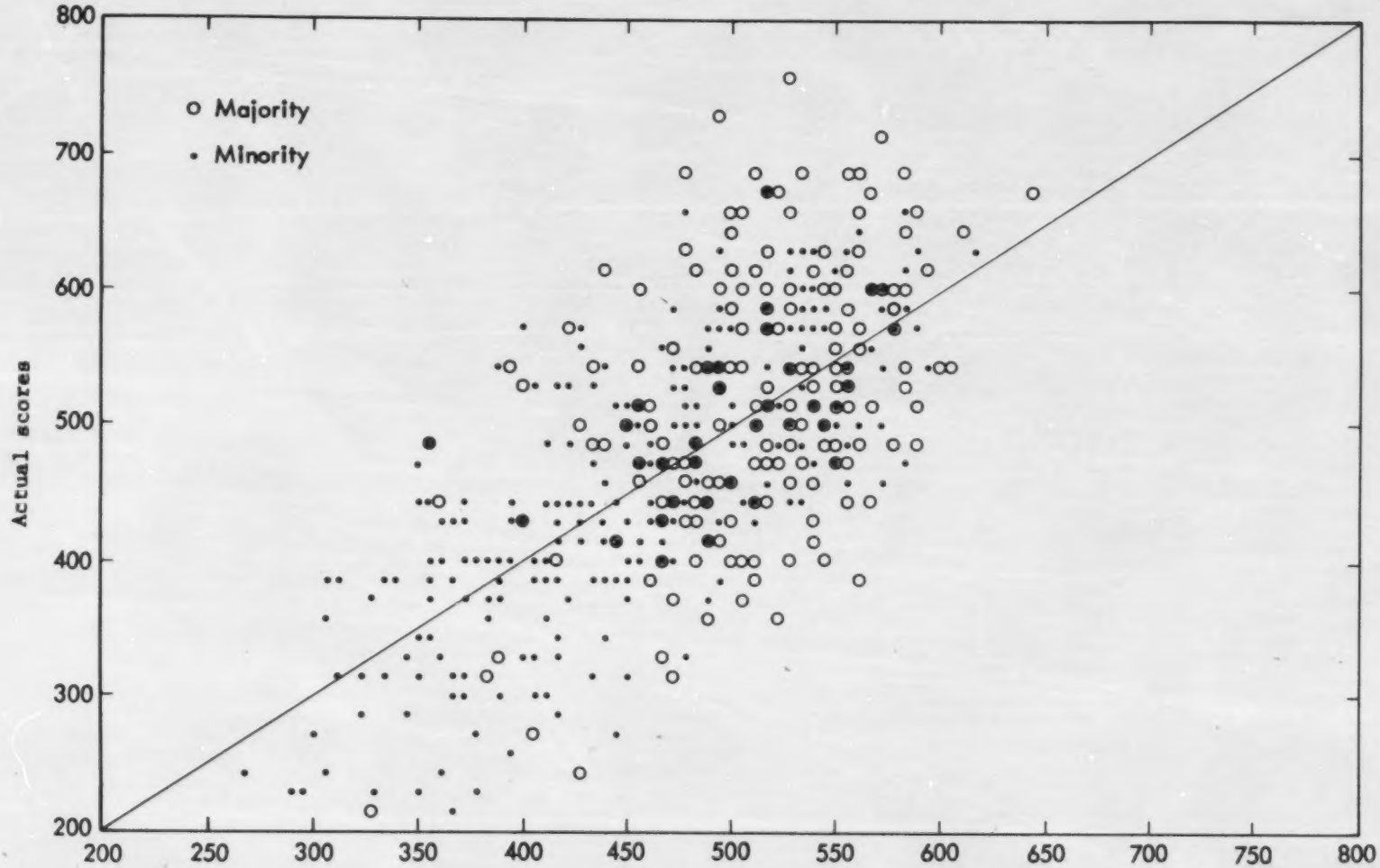

Fig. 3--Biochemistry scores: predicted vs. actual

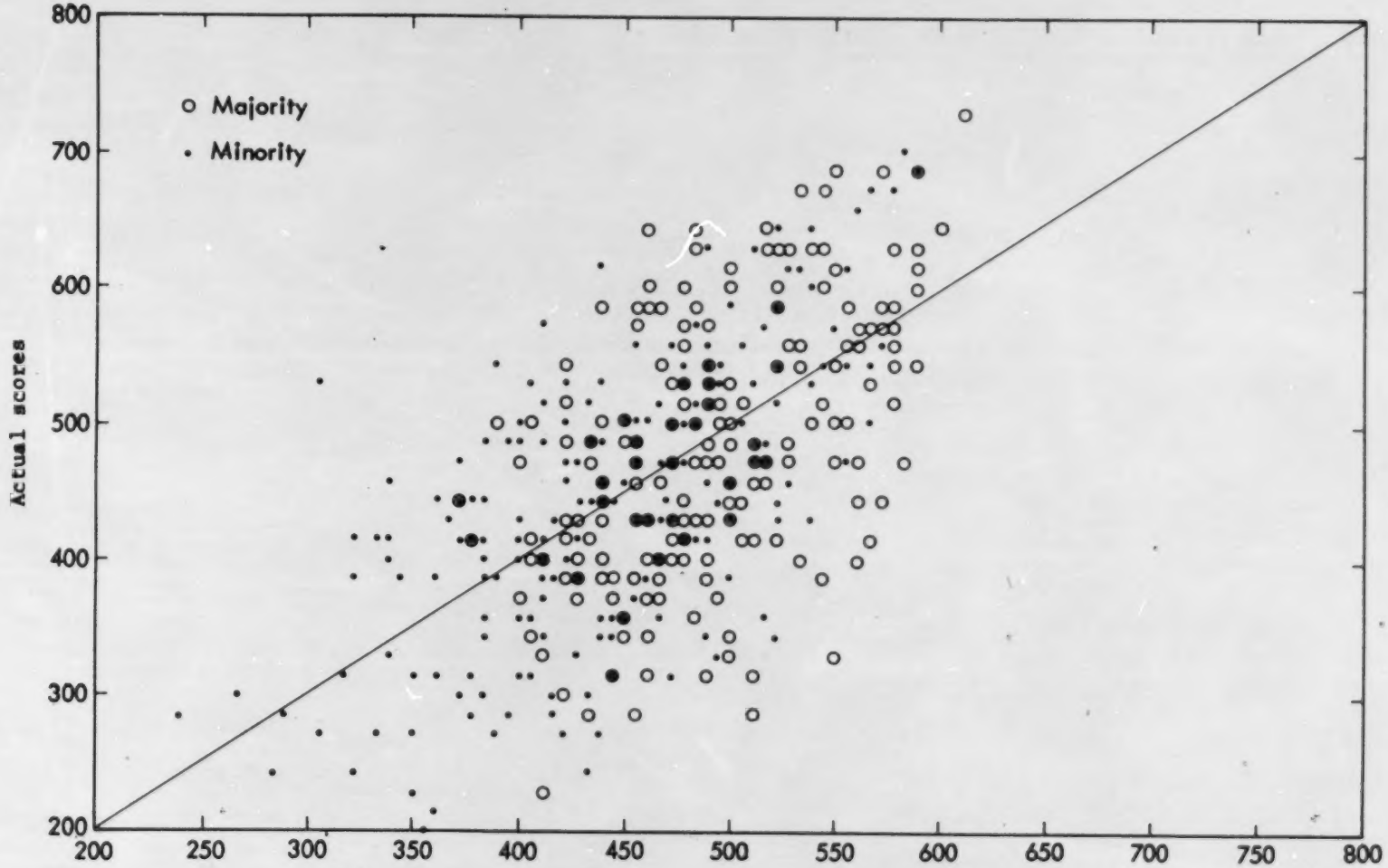

Fig. 4--Surgery scores: predicted vs. actual

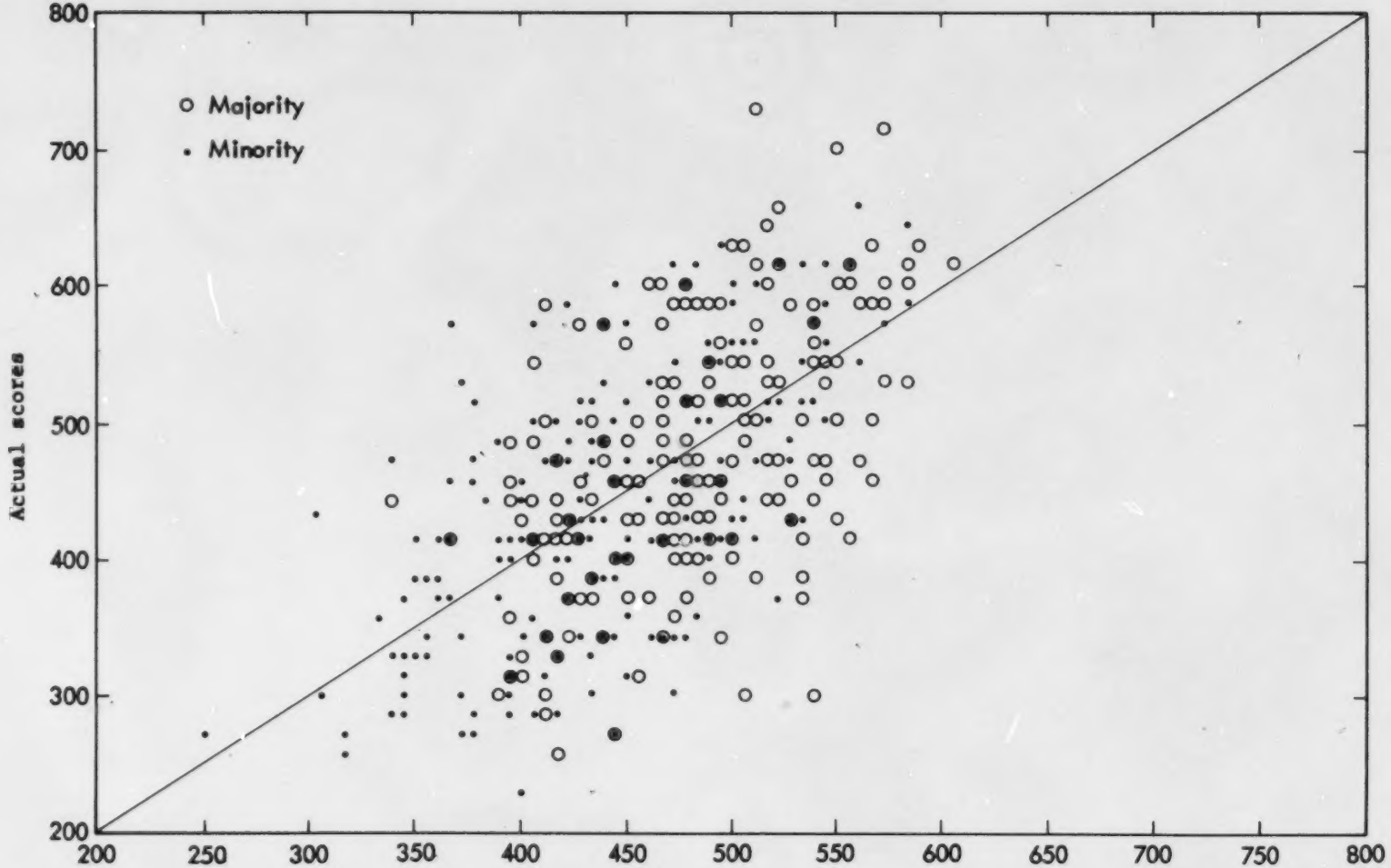

Fig. 5--Medicine scores: predicted vs. actual

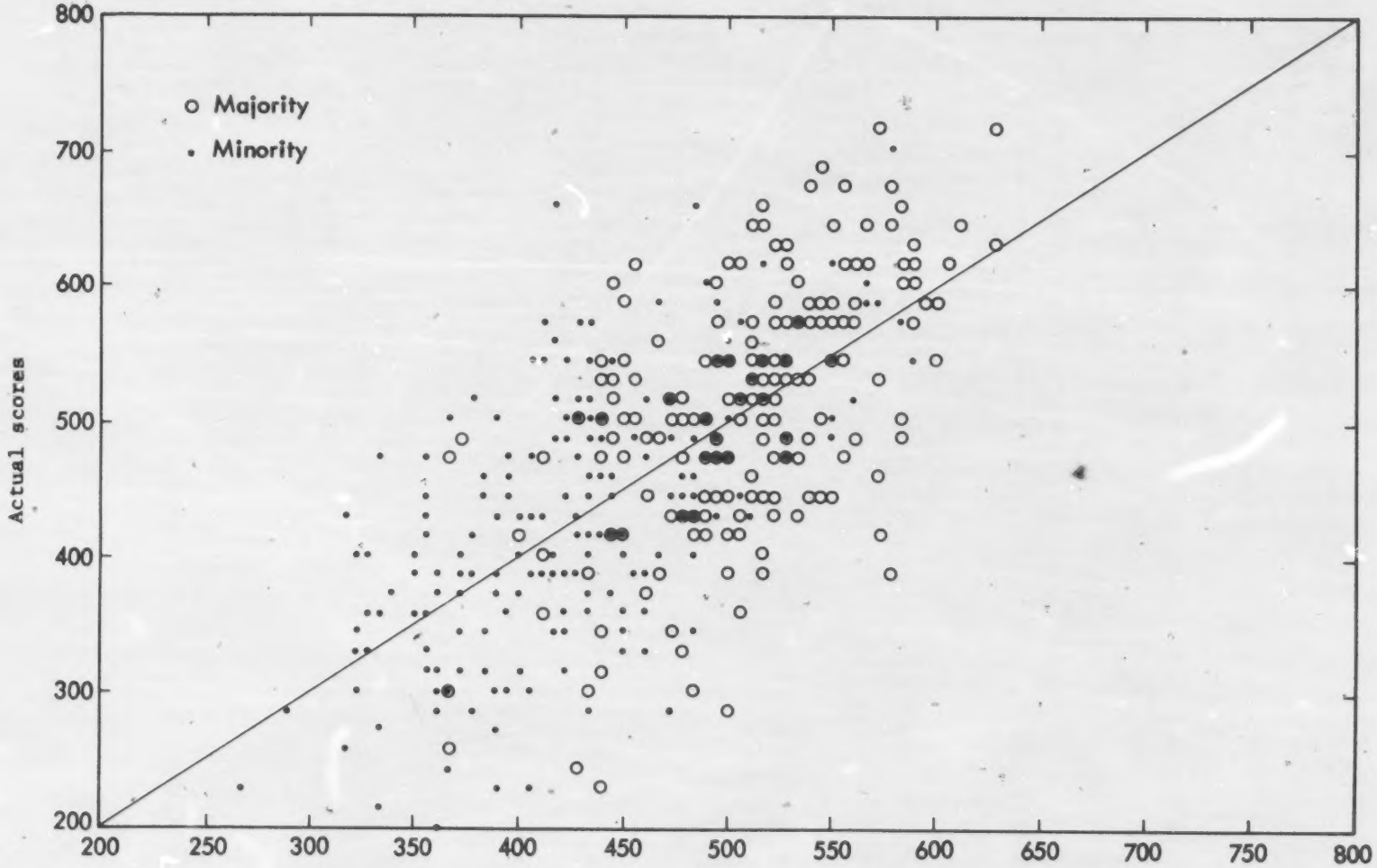

Fig. 6--Psychiatry scores: predicted vs. actual

tendency of the dots to be to the left of the circles reflects lower predicted scores for minority students. The larger vertical and horizontal spread of the minority points compared with majority points shows the greater homogeneity of the majority students' mean Part I scores and their preadmission characteristics. The minority points are closer, on average, to the diagonal than majority points because the minority regression has a smaller estimated standard deviation (69) than the majority regression (72). Figure 1 strikingly illustrates that although the minority average mean Part I score is 74 points below the majority average mean Part I score most of the minority student scores are contained within the majority student scores (over 400).

Figure 2 plots predicted and actual mean *Part II scores*. Its appearance is similar to the mean Part I plot although the minority-majority overlap is more substantial. Figures 3 through 6 show the predicted versus actual score plots for biochemistry, surgery, medicine, and psychiatry. As described in App. A, we chose these subsection scores for analysis because of their diversity. For example, the *range* of biochemistry scores for the 200 plotted majority students actually contains the minority student range. Psychiatry has the significantly fewer high minority scores (over 600) than the other subsections. The reader may use these plots to visually test his own hypotheses about the relative characteristics of these score distributions.

Although these plots give some useful comparative information about majority and minority students, they leave an important question unanswered. The predicted scores are calculated with different equations for minority and majority students. To compare how minority and majority students *with the same characteristics* will score, we must compare the two equations themselves.

ADJUSTMENT OF MINORITY AND MAJORITY PREDICTIONS FOR COMPARABILITY

All other things equal, how does minority student performance compare with majority student performance? The raw average scores are 74 points higher for majority students than for minority students on

both mean Part I and mean Part II. We need to control for the effect of a student's predictor characteristics when comparing scores. There are some 20 predictor variables. We use our regression equations to control for the effects of these variables to estimate the difference between the expected scores of a minority student and a majority student *who each have a given set of preadmission characteristics.* The estimated standard errors of the regression also allow us to compare *individual* scores of a random minority and majority student.

The regression equations for majority and minority students give predicted scores for any given set of characteristics. Since the regression equations have different coefficients, the differences between their predictions will vary with the values of the independent variables. For a student whose characteristics are those of an average minority student, the difference between majority and minority student mean Part I score predictions is 30 points. This difference is 24 points for a student whose characteristics are those of an average majority student. We will confine our comparative analysis to mean Part I and mean Part II scores. More generally we will "adjust" majority and minority predictions to the same population of predictor characteristics to compare them.

Mean Part I

Figure 7 presents the Part I predictions for majority and minority students adjusted to several sets of values of the predictor variables. The "minority student population" graph gives the majority and minority student regression equation predictions for the following values of the independent variables:

1. All four MCATs and undergraduate GPAs are (a) at the minority and (b) one half or one standard deviation (of the minority student distribution) away from the minority mean.
2. All other variables are at the minority mean.

The "majority student distribution" graph gives the majority and minority student regression equation for the comparable majority student values of the independent values. Table 6 tabulates these independent variables for the points in the majority and minority student populations

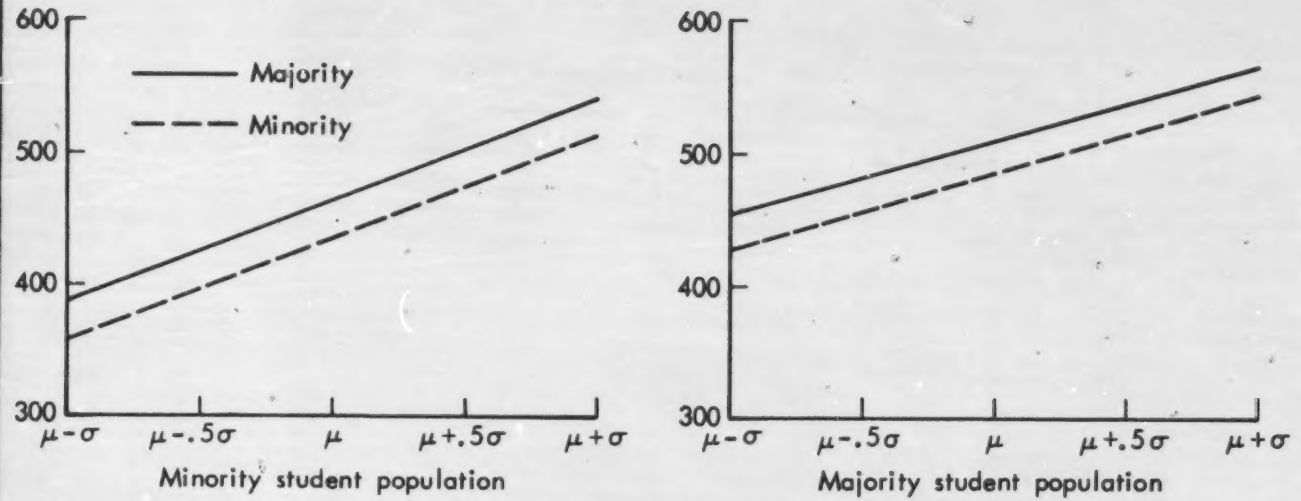

Fig. 7—Mean part I predictions

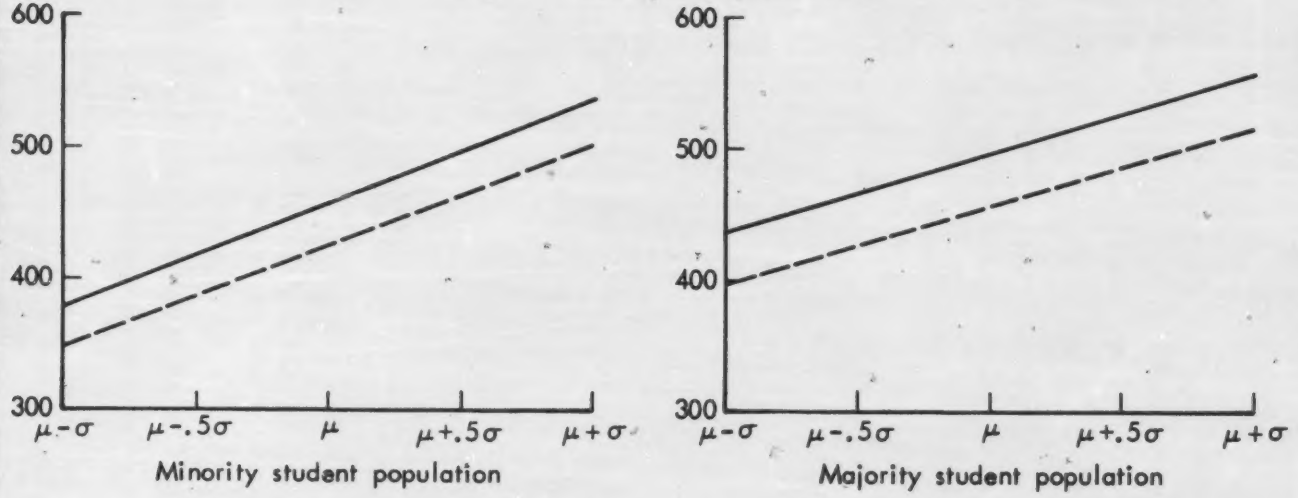

Fig. 8—Mean part II predictions

Table 6

DISTRIBUTION OF MINORITY AND MAJORITY STUDENT CHARACTERISTICS FOR PART I EXAM TAKERS

I. Characteristics That Vary[a]

	Mean $-\sigma$	Mean $-1/2\sigma$	Mean	Mean $+1/2\sigma$	Mean $+\sigma$
Verbal MCAT	498 (417)	538 (466)	578 (515)	617 (564)	657 (613)
General Information MCAT	498 (408)	534 (452)	570 (496)	606 (540)	641 (584)
Quantitative MCAT	563 (461)	598 (518)	633 (573)	668 (629)	703 (684)
Science MCAT	522 (433)	556 (479)	589 (526)	623 (572)	656 (619)
Undergraduate GPA	3.07 (2.62)	3.24 (2.87)	3.41 (3.11)	3.58 (3.36)	3.75 (3.60)

II. Characteristics That Do Not Vary

	Majority Mean	Minority Mean
School 0	.12	.03
School 1	.11	.05
School 2	.11	.22
School 3	.10	.27
School 4	.22	.08
School 6	.12	.15
School 7	.06	.09
School 8	.09	.05
School 9	.07	.06
Part I Exam Date	1.30	1.25
Missing Part I Exam Date	.17	.36
Year 1971	.37	.41
Repeat	.03	.05
Female	.15	.18
Selectivity Index	.31	.30

[a]Values in parentheses are for minority students.

as described above. For example, a student whose characteristics are half of a minority standard deviation below the minority mean will have an undergraduate GPA of 2.87 and MCAT scores of 466, 452, 518, and 479. The predicted scores graphed in Fig. 7 are tabulated in Table 7, where the majority equation prediction for such a student is 396 and the minority equation prediction is 427, a difference of 31 points.

Are the observed differences between majority and minority student predictions "real" or are they merely statistical artifacts? The regression equations give a prediction of the *average score of a majority (or minority) student* with a certain set of predictor characteristics. Since the regression equations themselves are estimated from the data, each coefficient has a standard deviation associated with its estimate. (See Table 4 and Table 5 for these values.) Thus, the prediction itself has a standard deviation, given in Table 7. The squared standard deviation (or variance) is given in Eq. (C.9), Appendix C. The difference between average minority and majority student performance on mean Part I for any of the given sets of predictor characteristics is statistically significant.[1] That is, the observed difference of between 22 and 32 points in predicted performance is inconsistent with its being a statistical artifact.

Our finding that, on average, minority students do get lower mean Part I scores than their majority student colleagues whose other characteristics are the same says very little about how scores of *individual* minority and majority students compare. The mean Part I minority and majority regressions have standard deviations about the regression line of 69 points and 72 points, respectively. Since the difference is never greater than 32 points, it is clear that the distributions of the scores of comparable majority and minority students have considerable overlap. A useful measure of comparing these two distributions is *the probability that a randomly selected majority student's Part I score will exceed a randomly selected comparable minority student's Part I score.* If the two distributions were the same, this probability would be .5. For each set of predictor variable characteristics, the bottom line of Table 7 gives this

Table 7

COMPARABLE MINORITY AND MAJORITY STUDENT PREDICTIONS AND ASSOCIATED ERRORS FOR MEAN PART I

	Minority Student Population[a]					Majority Student Population[a]				
	$\mu-\sigma$	$\mu-.5\sigma$	μ	$\mu+.5\sigma$	$\mu+\sigma$	$\mu-\sigma$	$\mu+.5\sigma$	μ	$\mu+.5\sigma$	$\mu+\sigma$
Minority Prediction	356.71	395.99	435.26	474.53	513.12	425.58	454.76	483.96	513.15	542.35
(Prediction Standard Deviation)	(6.57)	(5.02)	(4.59)	(5.55)	(7.37)	(7.53)	(7.76)	(8.35)	(9.25)	(10.37)
Majority Prediction	388.07	426.83	465.60	504.37	543.14	451.58	479.84	508.10	536.36	564.61
(Prediction Standard Deviation)	(5.44)	(4.19)	(3.11)	(2.47)	(2.60)	(2.31)	(1.86)	(1.78)	(2.12)	(2.72)
Estimated Probability that a Majority Student Score is greater than a Minority Student Score	.62	.62	.62	.62	.62	.60	.60	.59	.59	.59

[a] The symbols "μ" and "σ" denote the mean and standard deviation of the population.

probability.[2] Although the minority regression line is below the majority regression line by a statistically significant amount, the probability that the score of an individual majority student would exceed that of a comparable minority student *never exceeds 62 percent.* This probability varies between .59 and .62 depending on the pre-admission characteristics of the student. Thus, as our earlier plots indicated, majority students have a moderate edge in their chances of doing better than their minority counterparts in mean Part I scores.

Mean Part II Scores

Table 8 shows the distribution of student characteristics for the Part II analysis and corresponds to Table 6 for the Part I analysis. Table 9 presents the mean Part II score predictions for minority and majority students adjusted to the ten sets of values of the predictor variables. The predicted scores are graphed in Fig. 8. The differences between the predicted mean Part II scores of majority and minority students with particular sets of predictor characteristics range between 36 and 43 points. The comparable range for mean Part I prediction differences was smaller (22 to 31 points). As was the case with mean Part I predictions, the difference between *average* majority student performance and *average* minority student performance on mean Part II is statistically significant for any of the given sets of predictor characteristics.

In a comparison of *individual* performance of comparable majority and minority students, the probability that a random majority student's mean Part II score will exceed a comparable random minority student's mean Part II score is given on the bottom line of Table 9. These probabilities range from .64 to .66 depending on the characteristics of the students being compared--a moderate edge for majority students.

The discrepancy between comparable majority and minority student performance is larger for mean Part II scores than for mean Part I scores whether we are comparing average or individual performances. For both mean Part I and mean Part II, the observed difference between *average* performance of comparable majority and minority students is

Table 8

DISTRIBUTION OF MINORITY AND MAJORITY STUDENT CHARACTERISTICS FOR PART II EXAM TAKERS

I. Characteristics That Vary[a]

	Mean $-\sigma$	Mean $-1/2\sigma$	Mean	Mean $+1/2\sigma$	Mean $+\sigma$
Verbal MCAT	501 (426)	541 (474)	580 (523)	619 (571)	659 (620)
General Information MCAT	499 (415)	535 (458)	571 (501)	607 (544)	643 (588)
Quantitative MCAT	564 (480)	599 (533)	634 (586)	669 (639)	704 (692)
Science MCAT	520 (444)	553 (489)	586 (533)	620 (577)	653 (621)
Undergraduate GPA	3.05 (2.66)	3.22 (2.90)	3.39 (3.14)	3.56 (3.38)	3.73 (3.62)

II. Characteristics That Do Not Vary

	Majority Mean	Minority Mean
School 0	.08	.02
School 1	.11	.03
School 2	.12	.22
School 3	.11	.29
School 4	.15	.05
School 6	.16	.18
School 7	.07	.10
School 8	.10	.05
School 9	.10	.08
Part II Exam Date	2.65	2.67
Missing Part II Exam Date	.07	.12
Year 1971	.47	.49
Repeat	.02	.03
Female	.15	.17
Selectivity Index	.33	.32

[a]Values in parentheses are for minority students.

Table 9

COMPARABLE MINORITY AND MAJORITY STUDENT PREDICTIONS AND ASSOCIATED ERRORS FOR MEAN PART II

	Minority Student Population[a]					Majority Student Population[a]				
	$\mu-\sigma$	$\mu-.5\sigma$	μ	$\mu+.5\sigma$	$\mu+\sigma$	$\mu-\sigma$	$\mu-.5\sigma$	μ	$\mu+.5\sigma$	$\mu+\sigma$
Minority Prediction	343.67	383.01	422.35	461.72	501.10	395.94	425.68	455.42	485.16	514.90
(Prediction Standard Deviation)	(7.58)	(5.82)	(5.29)	(6.32)	(8.34)	(8.80)	(9.05)	(9.75)	(10.81)	(12.13)
Majority Prediction	379.64	419.22	458.79	498.37	537.94	437.17	467.27	497.37	527.46	557.56
(Prediction Standard Deviation)	(6.08)	(4.73)	(3.61)	(2.98)	(3.16)	(2.75)	(2.32)	(2.29)	(2.68)	(3.33)
Estimated Probability that a Majority Student Score is greater than a Minority Student Score	.64	.64	.64	.64	.64	.66	.66	.66	.66	.66

[a]The symbols "μ" and "σ" denote the mean and standard deviation of the population.

different from zero by a statistically significant amount--the difference appears to be "real." By contrast, an *individual* majority student never has better than a "2 to 1" odds in his favor of scoring higher than a comparable individual minority student in either mean Part I or mean Part II.

A SINGLE EQUATION FOR MAJORITY AND MINORITY STUDENTS

Examination of the corresponding majority and minority regression coefficients in Sec. III together with the exploratory analysis described in Sec. III and App. B led to two different regression equation specifications for predicting mean Part I and Part II scores for *all* students. We present these single regression equations for mean Part I and mean Part II as an alternative to our earlier separate equation analysis for comparing majority and minority student performance. This should be viewed as a simple, approximate alternative method.

Mean Part I

Table 10 presents regression equations fitted to all students' mean Part I and Part II scores. As might be expected, if the combined equation is appropriate, the combined equation standard deviation for mean Part I about the regression line is between the majority student regression standard deviation of 72.1 and the minority student regression standard deviation of 68.7. The difference between majority and minority students in how preadmission characteristics affect performance is captured by different coefficients on four variables: minority dummy, verbal MCAT, quantitative MCAT, and science MCAT. From the coefficients and associated standard deviations given in Table 10, this list could be cut down to science MCAT and quantitative MCAT without appreciable loss of predictive power.

Because 88 percent (1,898 out of 2,165) of the combined Part I sample are majority students, the combined equation coefficients for mean Part I are virtually identical to the majority equation coefficients except for the combined equation separate minority effects coefficients, which are about the same as the corresponding minority equation coefficients. The prediction lines and individual performance

Table 10

RESULTS OF REGRESSION ANALYSIS FOR COMBINED MAJORITY AND MINORITY STUDENT PERFORMANCE ON NATIONAL BOARD EXAMS
(Standard deviations in parentheses)

Variables	Part I Mean	Part II Mean
Constant	28.242	-85.923[c]
	(26.595)	(33.360)
Minority	-24.325	6.453
	(35.079)	(38.219)
School 0	9.032	17.862
	(7.966)	(11.431)
School 1	-42.824[c]	-36.270[c]
	(92.382)	(9.538)
School 2	42.391[c]	21.973[b]
	(11.811)	(8.775)
School 3	-2.215	15.767
	(92.663)	(10.905)
School 4	-13.313	6.198
	(88.708)	(8.357)
School 6	-1.919	-34.442[c]
	(83.182)	(8.813)
School 7	-12.097	-28.630[c]
	(88.238)	(9.747)
School 8	10.554	-19.182[b]
	(82.818)	(9.195)
Exam date	-11.149[c]	19.541[c]
	(41.650)	(7.047)
Missing Exam Date	-3.207	60.331[b]
	(10.945)	(24.597)
Majority Verbal MCAT[d]	.096[c]	
	(.027)	.135[c]
Minority Verbal MCAT[d]	.100[a]	(.033)
	(.059)	
General Information MCAT	.060[b]	.153[c]
	(.028)	(.035)
Majority Quantitative MCAT[d]	.004	
	(.027)	.086[c]
Minority Quantitative MCAT[d]	.177[c]	(.030)
	(.054)	
Majority Science MCAT[d]	.436[c]	.344[c]
	(.028)	(.034)
Minority Science MCAT[d]	.265[c]	.267[c]
	(.069)	(.066)
Undergraduate GPA	35.280[c]	32.081[c]
	(4.277)	(4.967)
Missing GPA	115.68[c]	114.940[c]
	(15.921)	(18.255)
Year 1975	10.808[c]	-7.985[a]
	(3.828)	(4.791)
Repeat	-6.541	2.614
	(9.489)	(13.146)
Female	-17.956[c]	2.742
	(4.398)	(5.497)
Selectivity Index	25.616[c]	20.204[b]
	(7.154)	(8.356)

[a] Significant at the .10 level.

[b] Significant at the .05 level.

[c] Significant at the .01 level.

[d] Differences between majority and minority student performance in the effects of preadmission characteristics are captured by different coefficients on these variables. For the Part II mean prediction, the effects of Verbal MCAT and Quantitative MCAT were so similar that we chose to use a single coefficient for majority and minority students.

comparison for the combined equation should be almost the same as those in Fig. 7 and Table 7. The combined equation gives a somewhat more compact way of making comparisons.

Mean Part II Scores

Table 10 presents the fitted regression equation to all students' mean Part II scores. This equation is comparable to the separate majority and minority student mean Part II equations as described above for the mean Part I comparison, and the conclusions are the same as well.

Footnotes--Section IV

[1]Equation (C.11) gives the formulas for testing whether the two means are equal. In all cases, a two-sided test that the difference is zero is rejected at the .01 level.

[2]See Eq. (C.13) and the related discussion for the method of computing these probabilities.

V. CONCLUSIONS

The objective of this report has been to improve prediction of minority and majority student performance in the medical school curriculum. We are more concerned with minority than majority student performance because medical schools admissions committees have encountered difficulty in selecting minority applicants who progress satisfactorily through the medical curriculum. By contrast, the problem for admissions committees screening majority applicants is to decide whom to admit from the large number who could progress satisfactorily.

The subjects for our analysis were members of the graduating classes of 1975 and 1976 from nine medical schools. The performance measures are scores on the Part I and Part II examinations prepared by the National Board of Medical Examiners. The data we used are available to every medical school: Medical College Admission Test scores, grade point average, and other descriptive data that we have quantified so as to make them usable in statistical analysis.

Our analysis is comprehensive when compared with other studies dealing with minority and majority student performance, and it has important limitations. The NBME Part I and Part II exams are the most widely used standardized tests for assessing medical school achievement, but they do not attempt to measure some important attributes of good medical practice (e.g., manual skills, rapport with patients, professional diligence). Our prediction models exclude some sources of information that may be important (e.g., assessments based on applicant interviews, letters of recommendation, nonacademic achievements). However, because we restrict our analysis to standardized measures, we believe the results can be interpreted unambiguously for any medical school.

On average, minority students rank below majority students on the standardized measures used by admissions committees to assess an applicant's preparation for medical school. Various data have been published quantifing some of the minority-majority differences. The

significant differences are sufficient to correctly infer that minority students, as a group, will score below majority students on the NBME exams.

The objectives of affirmative action programs relate to minority groups that are underrepresented in professions such as medicine, but it is decisions on individuals that drive those programs. We have therefore focused our analysis on the problem of predicting how an *individual* student--minority or majority--will perform on the basis of premedical school characteristics and the medical school attended.

PREDICTION ACCURACY

Our analysis shows that medical schools should be able to predict the National Board exam scores of their students with reasonable accuracy on the basis of preadmission characteristics. If the nine medical schools in our sample had used our models to predict Part I and II mean scores for majority students in their 1975 and 1976 graduating classes, they would have been correct within 100 points for about five out of every six students. In similar minority student predictions, they would have been correct within 100 points for about seven out of every eight students. The accuracy is only slightly less for some individual parts of the exams. We fitted equations to four individual parts of the exams in addition to the Part I and Part II means scores.

We achieved the greater prediction accuracy for minority students than for majority students by weighting their preadmission characteristics differently. In particular, we observed that the quantitative and verbal parts of the MCAT exam have considerable predictive power for minority students. Although the science MCAT and undergraduate GPA are the best predictors of majority student performance in all but one of our fitted equations, verbal MCAT has better predictive power than either in all but one of the minority equations. The quantitative MCAT is superior to the science MCAT in one of the minority equations and superior to GPA in three, suggesting that measures of general aptitude may be better predictors of performance than measures of past achievement, within the ranges observed for minority medical students.

Another notable difference between minority and majority students is the effect of the undergraduate school attended. The selectivity of the undergraduate school has a statistically significant positive effect in four of the six majority performance equations. That variable never approaches a level of statistical significance in any of the six minority equations.

We cannot predict either majority or minority medical student performance on the NBME exam very accurately without knowing which medical school they attended. School-to-school differences in performance appear to be of two types. In one, a consistent positive or negative school effect appears across all exam parts; it may be caused by the differences in the importance a school attaches to the NBME exam as an evaluation tool. In the other type, there are significant differences across schools but the effect is positive for some sections of the exam and negative for others. These differences may be more properly attributable to characteristics of parts of the curriculum than to the school as a whole.

The medical school effects on NBME exam scores for minority and majority students are similar, although there is weak evidence that both positive and negative school (or curriculum) effects on majority student performance may be amplified in the case of minority students. However, neither the pattern nor the strength of school effects on minority performance indicates that one or another institution has a superior program for helping minority students to compensate for deficiencies in their academic preparation.

UNEXPLAINABLE DIFFERENCES IN PERFORMANCE

Although the prediction accuracy of our models is sufficient to assist in the assessment of medical school applicants, substantial differences in performance across individuals are not explained by standard preadmission measures of academic preparation or by medical school effects. There are two kinds of differences in performance: One is the expected difference between actual and predicted performance for individuals, reflecting the limitations of our models. The other is the difference between expected performance of a minority and a majority student having the same preadmission characteristics.

Differences between predicted and actual performance are inevitable in any model. We would expect these prediction errors to occur for some combination of four reasons. One is the *structural limitations* of our simple linear model. Although this model was superior to any other we tried, it may fail to capture some complex relationships between the predictor variables and NBME exam performance. Another source of prediction error is *measurement error in the predictor variables* it uses--the difference between an individual "true score" on a test (or the GPA), which reflects true competence in the tested field, and the recorded score for that evaluation. The third reason for error is *omitted variables*. The models do not take into account a number of individual characteristics that, if available, might permit more accurate prediction (e.g., dedication to a medical career, psychological attributes). The fourth source is *measurement error in the NBME exam*. That is, the NBME exam measures "true ability" with error; a perfect predictor of true ability is not a perfect predictor of NBME scores.

Our analysis indicated that on average a majority and a minority student with otherwise equivalent premedical school characteristics will not perform equally well on the NBME exams. In such circumstances, we should expect the majority student to score higher than the minority student in about two out of three cases. We do not know the reason for this difference, but it is too consistent across models and student types to be dismissed as a statistical artifact.

We offer several possible explanations, but only as hypotheses worthy of testing. There could be *omitted variables* in our models as mentioned above. This could be a partial explanation for the size of the differences in performance we see between equivalent minority and majority students.[1] There may be greater cultural or other effect in the NBME exam than in the predictor variables, and therefore minority scores substantially understate knowledge in the various areas tested. Minority undergraduate grade point averages may be inflated and give falsely high expectations of performance if judged by majority standards.[2] Admissions committees may fail to take into account factors not in our prediction models that would enable them to predict minority student performance more accurately. Exploitation of such

factors in admissions decisions on majority applicants might explain a difference in performance of the sort we observed.

IMPLICATIONS FOR THE MEDICAL SCHOOL ADMISSIONS PROCESS

Our analysis has implications for the medical school admissions process to the extent that expectations about an applicant's eventual performance on the NBME exams are judged as important to the admissions decision. We know there are large differences across schools in the importance attached to the NBME exams, but we are confident that few if any schools would regard the exams as irrelevant to student evaluation.

Whether a medical school uses the full range of possible scores of the NBME exams to evaluate performance or uses a threshold of satisfactory performance, our models can help admissions committees make more confident decisions. Most committees may feel that the quality of the applicant pool is sufficiently high to make the benefits of more confident predictions of majority student *academic* performance of little value. By contrast, there is widespread interest in finding better means for predicting minority student performance. The committees' concerns are twofold. They want to avoid admitting a student who will be unable to complete the curriculum satisfactorily. They also want to avoid rejecting an applicant who displays the qualities to become a good physician but who is incorrectly judged incapable of satisfactory academic performance.

It is not surprising that admissions committees feel they have inadequate bases for predicting minority applicant performance because their experience is limited to observations of few minority students over the recent period of affirmative action programs. The precision of our prediction models can be attributed mainly to the benefits of the larger population of minority students from nine schools. Medical schools can obtain similar prediction accuracy by centralized analysis of pooled data on minority student attributes and performance.

A comparison of results of our analysis of minority student performance with the results of analysis of the minority student admissions process suggests that admissions committees might well wish to

reevaluate criteria they now use. All nine schools heavily emphasize measures of undergraduate achievement in science (science GPA and science MCAT) in admission decisions.[3] However, only two of the nine assigned significant weight to the more general aptitude measures (the verbal and quantitative MCATs), which are of comparable importance to the accuracy of performance predictions.[4]

Our comparative analysis of minority and majority student performance suggests that admissions committees should seek to improve their capacity to evaluate the *nonquantifiable* attributes of minority applicants. Deficiencies in this area are one possible explanation for the observed differences in performance among minority and majority students whose *quantifiable* attributes are equivalent.

Admissions committees are now in the process of making decisions on the basis of a completely revised Medical College Admission Test.[5] This new MCAT is designed to overcome some deficiencies of its predecessor, on which our analysis is based; one important consideration in the revision was to improve the assessment of minority applicants. It is too soon to know either how admissions committees will use the new test or how well it will serve as a predictor of student performance. However, we believe there are enough similarities between some components of the old and new MCAT that the results of our analysis can suggest ways to use the new instrument. We also believe that the analytic approach we have taken is quite well suited to early evaluation of the new MCAT.

More important than new admissions criteria are better measures of *practicing physician performance*. Without such measures, analysis of the sort we have conducted cannot be wholly satisfying. However, analysis using admittedly deficient evaluation instruments (MCATs, GPA, NBME exams, etc.) remains relevant to the kinds of physicians who treat us so long as those instruments play an important role in the admission to and the assessment of performance in medical school.

BROADER IMPLICATIONS

In the recent Bakke decision, the Supreme Court held that medical schools, and the higher education community, more generally, need not

be "color blind" when evaluating applicants for admission to their programs. The thurst of the opinion by Justice Powell is that special consideration, i.e., racial preferences, may be accorded minority applicants in the interests of social justice. Our research does not suggest what amount of preference is appropriate for affirmative action programs. Rather, it indicates that, quite apart from consideration of social justice, medical school admissions committees' predictions of applicant performance will be inferior if they are based on a "color blind" assessment of commonly used achievement and aptitude measures.

Our research deals only with the problems of predicting medical school performance. However, the findings suggest that empirical analysis of a similar type might help other professional schools improve their affirmative action programs, in the sense of selecting the most promising minority applicants.

Footnotes--Section V

[1]To see how this might occur, suppose that there is an additional omitted variable, x_A (say "socioeconomic status"), whose partial correlation with race (1 = majority, 0 = minority) is positive given the other explanatory variables in the model. Then, our estimate of the difference in performance due to race would be biased upward. See A. S. Goldberger, *Econometric Theory*, Wiley, New York, 1964, pp. 194-196, for the exact formula. A number of other studies of achievement or performance that control for individual characteristics find statistically significant differences between minority and majority outcomes. See Z. Griliches and W. H. Mason, "Education, Income and Ability," *Journal of Political Economy*, 80, No. 3, Part II, May/June 1972, pp. S74-S103; and W. L. Hanson, B. A. Weisbrod, and W. J. Scalon, "Schooling and Earnings of Low Achievers," *American Economic Review*, 40, No. 3, June 1970, pp. 409-418.

[2]At most this explains only part of the difference. To explain the full observed difference, minority grades would have to be overstated by one full gradepoint. That is, the average minority student whose average was a high B would have earned only a high C.

[3]These admissions models are reported in Williams et al., *Factors Affecting Medical School Admission Decisions for Minority and Non-Minority Applicants: A Comparative Study of Ten Schools*, R-2030-HEW, The Rand Corporation, forthcoming 1978.

[4]Since the performance analysis used data on students (admicted applicants) and the admissions analysis used data on both admitted and nonadmitted applicants, the conclusions of the two analyses are about different populations. Our statistical models control for this difference under the assumption that the relationship between the predictor variables and the NBME scores would be the same for nonadmitted applicants as for students who took the exam. This assumption cannot be verified using our data.

[5]The new MCAT is described in Mary Littlemeyer (ed.), *The New MCAT Student Manual*, Association of American Medical Colleges, Washington, D.C., 1976; and Division of Educational Research and Measurement, *New Medical College Admission Test: Interpretative Manual*, Association of American Medical Colleges, Washington, D.C., 1977.

Appendix A
DATA AND VARIABLES

This appendix gives details on the data collection, data limitations, and the variables created from the records of the ten schools. Data collection is always a difficult task, and collecting data from several different sources is even more difficult. We experienced difficulties in this study because schools keep records in formats that are useful to them and for their own purposes. As purposes change, so do the form and the content of student records.

The data for this study were drawn from the student records of ten medical schools. To protect the privacy of the student data, we assigned each student a code number, and the schools kept the key associating the code numbers with student names. We restricted our analysis to variables common to all schools' records, or whose values could easily be estimated or readily calculated from other data in a student's record (e.g., undergraduate GPA from science GPA, nonscience GPA and the units of each). The data on premedical characteristics were four parts of the Medical College Admission Test scores, undergraduate GPAs, undergraduate school attended, sex, and race. Data on medical school characteristics were class to which the student was admitted, and date of taking Part I and Part II of the National Board of Medical Examiners examination. Data on medical school performance are the results of Part I and Part II of the NBME examination as well as whether the student repeated at least one year of medical school.

Since all MCAT scores were used in the analysis of Part I and Part II performance, any student record that did not have all four MCAT scores was deleted from the analysis. The MCAT scores were so central to the analysis that we decided not to estimate the average MCAT effect of missing values as was done for some other variables. Fortunately, few MCAT scores were missing. Students who transfer from another medical school may have incomplete records at the new school; sometimes students who had taken more than one MCAT examination had no score recorded, and there were the usual coding errors.

Undergraduate GPAs are recorded on a four point scale by most medical schools (A=4, B=3, C=2, D=1, F=0). The American Medical College Application Service (AMCAS) calculates grade point averages for individual students and reports them cumulatively and separately for each year to subscribing schools. AMCAS separates biology, chemistry, physics, and mathematics into one category and calculates this as a science GPA; all other grades are averaged in a nonscience category. Where available, we used the AMCAS-calculated cumulative undergraduate grade point averages. Our preference was to use science and nonscience grades separately, but we were unable to do so because two medical schools recorded only undergraduate GPA in their student records. Where undergraduate GPA was not available, it was calculated from data recorded for science and nonscience graded credit hours. If an overall GPA was not recorded and either science or nonscience were also not recorded, we calculated an overall GPA from the individual student's transcript.

After the various methods of acquiring overall GPAs for students were used, a nontrivial number of students remained who were missing GPAs. Rather than drop these records from the analysis, we included these students in the analysis by using a dummy variable when undergraduate GPA was missing.

Data on student's undergraduate college were obtained from their records and coded into ten groups using the classification of *Barron's Profile of American Colleges and Universities* as described in the text.

Rather than estimate equations for each subscore of Part I and Part II exam, we limited our attention to a few tests. Some tests tend to be highly correlated with other tests. Total Part I and Part II scores were selected because they represent the student's performance on the range of examination areas. Additional tests were selected on grounds of their low correlation with each other. For Part I of the NBME, biochemistry and behavioral science were least correlated. On Part II, surgery, psychiatry, and internal medicine had low correlation.

The behavioral science test of Part I on the NBME posed a special problem. Before 1973 there was no behavioral science section in Part I

of the NBME. The majority of one class at one school and large numbers of one class at another school took Part I before the behavioral science test was part of the Part I examination. The large numbers of missing observations forced elimination of the analysis of the behavioral science test.

Fewer students had recorded scores for Part II of the NBME than for Part I for two reasons: (1) the timing of the data collection and (2) a large number of transfer students. At the time of our final data collection, some students had not taken Part II of the NBME examination. Transfer students were lost because many of them did not have MCAT scores as part of their student records. The uncollectable and unavailable data, combined with the normal attrition, made a difference of 590 between the Part I and Part II sample sizes.

As indicated above, the two parts of the NBME examination can be taken over a range of times. At some medical schools, students are given a choice; at others they are not. In the latter cases, an entire class will take the examination as a unit. There were various conflicting expectations about the performance of students who took the examination early and those who took the examinations later in their medical training. There is no consensus among medical school deans and professors regarding the effect on score of when a student takes the NBME examinations. On the one hand, self-selection could lead to the best prepared taking the examination early; on the other hand, the best prepared may put the examination off to attain the highest possible score. A problem is that the examinations come during the year in which the student is taking courses in the subjects that appear on the examination. To empirically account for the effect of taking the examination early or late, two new variables were created that expressed the difference in years between the dates of the students taking Part I and Part II and the date the student entered medical school. When either the date of entry into the medical school or the examination date was missing, a dummy variable was used to indicate that the elapsed time was not known.

In addition to data problems, we created or modified several variables to make the analysis richer and take advantage of the form of the data. Students who have to repeat a year were generally expected to do less well on the NBME examinations than students who were making normal progress. A dummy variable indicated whether a student had repeated a year or more of medical school.

The numbers of students of certain ethnic backgrounds (e.g., native American) were very low. The small sample size necessitated some arbitrary decisions about grouping students by ethnicity. Clearly majority students formed one group. Blacks and Asians formed two separate groups, and all others constituted a final group. The last group was primarily Mexican-American and Puerto Ricans. Student records that were missing the entry for race were assigned to the majority group. Any further breakdown of minorities into constituent groups would be confounded because, in our sample, minority students from different groups tended to be disproportionately represented at various schools. For all these reasons we decided to assess the differential performance of the two broad categories of minority and majority students.

Additional variables included a dummy indicating whether the student was a member of the class of 1976 or 1975, to test for a difference between the two classes. When a student record was missing a graduation date, we assumed that the student was in the class of 1976. A dummy variable for sex was added; the default value of this dummy variable was male.

Tables A-1 and A-2 give the means and standard deviations of the variables used in the analysis.

Table A-1

MAJORITY STUDENT CHARACTERISTICS
MEANS AND STANDARD DEVIATIONS

Variable	Mean	Standard Deviation
For Part I Analysis (N = 1,899)		
Undergraduate GPA	3.41	.34
MCAT Verbal	577	79
MCAT General Information	570	72
MCAT Quantitative	633	70
MCAT Science	589	67
Selectivity Index	.310	.071
Mean Part I	509	87
Biochemistry	5[illegible]	95
For Part II Analysis (N = 1,369)		
Undergraduate GPA	3.39	.34
MCAT Verbal	580	79
MCAT General Information	571	72
MCAT Quantitative	634	70
MCAT Science	586	67
Selectivity Index	.330	.077
Mean Part II	498	91
Medicine	482	94
Surgery	492	99
Psychiatry	509	91

Table A-2

MINORITY STUDENT CHARACTERISTICS
MEANS AND STANDARD DEVIATIONS

Variable	Mean	Standard Deviation
For Part I Analysis (N = 266)		
Undergraduate GPA	3.11	.49
MCAT Verbal	515	98
MCAT General Information	496	88
MCAT Quantitative	573	111
MCAT Science	526	93
Selectivity Index	.302	.059
Mean Part I	434	101
Biochemistry	455	99
For Part II Analysis (N = 199)		
Undergraduate GPA	3.14	.48
MCAT Verbal	523	97
MCAT General Information	501	87
MCAT Quantitative	586	106
MCAT Science	533	89
Selectivity Index	.317	.063
Mean Part II	424	96
Medicine	444	97
Surgery	440	104
Psychiatry	428	99

Appendix B
EXPLORATORY DATA ANALYSIS

We performed extensive exploratory analysis of the data in order to choose our majority and minority student regression equations for predicting student performance cn Part I and Part II of the National Boards.

SCHOOL-BY-SCHOOL ANALYSIS FOR MAJORITY STUDENTS

We began our exploratory analysis by fitting regression equations to the scores for each of several sections of Part I and Part II of the NBME exam for majority students, separately for each of our nine medical schools. Although our final model included all the schools in a single equation, the analysis that led to our choice of independent variables in the separate school equations carried over to the combined school equation. The primary goal of our school-by-school analysis was the proper specification of the independent variables in regression equations for predicting NBME scores. The independent or predictor variables were the students' entering characteristics, which included undergraduate grade point average; selectivity of the undergraduate college; Medical College Admission Test scores; sex; and other variables that correlate well with National Board scores as described in Sec. II. The same independent variables were used for each equation, but different equations corresponding to six different dependent variables (the scores on different sections) were run for each of our nine schools. These dependent variables were mean Part I score; mean Part II score; score on the biochemistry section of Part I; and scores on the psychiatry, medicine, and surgery sections of Part II.

For our initial exploratory analysis, we decided to focus on two schools, initially allowing more detailed analysis than was possible if all nine schools were used. Our approach was to develop a suitable model for the first school, then apply that model to the second school to check consistency across schools.

We aimed for a model that predicts the performance of applicants, perhaps in much the same way as an admissions committee. Having established a suitable model, we could then analyze the differences between students' actual and predicted scores. The literature on predicting medical student performance suggests a variety of complex relationships between quantitative admissions data (e.g., admissions test scores, undergraduate grades, undergraduate school) and performance in medical school as measured by standardized tests. Most empirical studies of student performance cannot clearly explain the performance-admissions relationship, so we must rely wholly on our data for a model specification.

We tried various specifications of independent variables in regression equations for predicting NBME scores. Each fitted equation yielded a residual for each student (observed score minus predicted score). These residuals were plotted against each independent variable and against the predicted score. This analysis of residuals can suggest transformations of the independent or dependent variable or the inclusion of interactions among the independent variables.[1]

In estimating the regression equations for Part I scores for the first school, we tried a number of equations in an effort to capture the complex relationships between a student's premedical school attributes and his performance. Plots of the residuals against each of the MCATs indicated that no transformation was needed. We tried a series of transformations of grade point averages to capture the grade effect. There was a slight tendency for an increase in undergraduate GPA to have more of an effect for lower rather than higher GPA, but this tendency was too small to justify adding a squared GPA as a variable. A coefficient of science concentration (as measured by proportion of science hours taken) was computed and entered in the regression equations. It was never statistically significant when a GPA variable was also in the equation.

To test for the possible effect of the caliber of a student's undergraduate college, the selectivity index of the applicant's undergraduate school was entered into the regression.[2] Several transformations of the selectivity index for predicting Part I scores

worked well for one school, but when the same model was tried on other schools the coefficient of the transformed selectivity index was not statistically significant.

Given the small size of the classes from our school population, we hesitated to reject the existence of a school effect of some sort. So we then tried the selectivity index developed by Rolph, Williams, and Lee.[3] Their index was estimated from regression equations to show for majority students what effect having graduated from an undergraduate college with a certain Barron's index value had on being admitted to medical school. For this report, the coefficients for the entering class of 1972 were rescaled for ease in interpretation to make them range between zero and one. See Table B-1.

Table B-1

ADMISSION SELECTIVITY INDEX

Barron's Selectivity Index		New Selectivity Variable
Unknown or unclassified	0	.1214
Not competitive	1	.1214
Less competitive	2	.204
Competitive	3	.115
Competitive +	4	0
Very competitive	5	.2414
Very competitive +	6	.3087
Highly competitive	7	.5529
Highly competitive +	8	.6739
Most competitive	9	1.0000

We tried other equations to test the interaction effect of grades and undergraduate college. No version of an interaction term for selectivity index and grade point average was of additional help. We also looked for possible interactive effects among the MCAT tests. None of several interaction specifications proved to be as good as simply entering each of the MCAT scores separately.

Having failed to find more complex relationships, we concluded with a simple model for Part I scores. It is linear in the following set of independent variables: all four MCATs, the time lapsed between entry and taking the exam, undergraduate GPA, selectivity index, dummy variables for repeating a year, being female, entering in 1971, and dummy variables to account for some missing values. A similar analysis was carried out for Part II and for the subsections of both Part I and Part II, and we reached the same conclusions.

COMBINING MAJORITY STUDENTS ACROSS SCHOOLS

Some similarities in the coefficients of the nine different school equations led us to pool all the students in an attempt to find a single regression equation that would predict exam scores for majority students at all nine schools. There is almost a continuum of possible equations. At one extreme, we could have assumed that a student's school makes no difference at all in predicting his exam score--that is, attributes of students for all nine medical schools have the same coefficient. In that case there would be 13 independent variables (including the constant term) and 13 coefficients. The other extreme is a separate equation for each of the nine schools as described above, giving a total of 9 x 13 = 117 coefficients.

Between these two extremes, we examined regression equations that have some explanatory variables with separate coefficients for each school with the remaining explanatory variables having the same coefficients for all schools. An example of such a model is separate dummy variables for each school and separate science MCAT variables for each school, with the remaining variables being common across schools. This particular example has 2 x 9 + 11 = 29 coefficients, substantially fewer than the 117 coefficients when each school has its own equation.

Where a prediction model is needed, parameters (i.e., coefficients) should be introduced sparingly so that there is maximum resolution for each parameter introduced.[4] Following this principle, we attempted to use separate school coefficients only for those variables where the additional predictive power was sufficient to justify the trouble.

As a first step we compared the nine school equations for predicting Part I scores and the corresponding nine equations for Part II. For reasons of economy and manageability, we confined our exploratory analysis to the mean Part I and mean Part II scores and assumed that the form of the equation for a particular part applied to its sections (i.e., medicine gets the same form of equation as Part II).

Comparison of the coefficients of the nine separate school equations where there were *no separate school effects on any variables* made it clear that we could use a single variable with no separate school effects for both Part I and Part II for at least the following attributes: verbal MCAT, female, repeat, year, missing GPA, exam date, missing exam date. But separate school effects for the remaining attributes--science MCAT, quantitative MCAT, general information MCAT, undergraduate GPA, selectivity index, and school dummies--yields many more variables (61) than necessary. We tried a number of specifications by choosing which variables were allowed separate school effects and which were not. For Part I, we found that *only one variable was needed to capture the separate school effects, and the best one (highest R^2) was the school dummy variable.*[5] The two best variables to use with separate school effects were science MCAT and the school dummies. The same results held for Part II, except that the two best variables were selectivity index and the school dummies.

Section III reports the results of the Part I and Part II regressions as well as the subsection regressions--biochemistry, medicine, surgery, psychiatry--for the equations using 21 variables: nine school dummies with the 12 other independent variables being the same for all nine schools.

MINORITY STUDENTS

There are 268 minority students who took Part I and 199 who took Part II in our sample. The corresponding sample sizes of majority students are 1,899 and 1,309. This small number of minority students means that we could explore only fairly simple structures for predicting their performance. As a result, the *structure* (but

not the coefficients) of our final minority score prediction equation relies as much on the majority student exploratory data analysis as it does on that of minority students.

No meaningful school-by-school exploratory analysis of minority student scores was possible because of the small number of minority students at each school. (The 268 Part I NBME minority exam takers were spread across the nine schools as follows: 8, 13, 59, 72, 22, 40, 24, 14, 16.) However, by combining minority students across schools, we were able to examine which variables required interaction terms with each of the nine schools. Our basic independent variables were the same ones used to predict majority student scores. We begin with the minority regression equation for Part I and Part II scores using *no separate school effects on any variables*. By including school effects in different ways, the power of the equation could be substantially increased. The final equation for predicting minority scores included a dummy variable to capture the school effect and a single value for all other variables across the nine schools. The exploratory analysis leading to this decision included looking for separate school effects for the following variables by fitting the appropriate regression equations: school dummy variables, science MCAT, quantitative MCAT, and selectivity index. We fitted equations including only one variable each time with a separate school effect (giving 9 + 12 = 21 variables).

We found that using two separate school effect variables led to a regression equation with 2 x 9 + 11 = 29 coefficients, too many for reliable interpretation with our minority student sample sizes. We therefore decided to use school dummies rather than other separate school variables, although a separate science MCAT variable equation had almost as much explanatory power.

A SINGLE EQUATION FOR MAJORITY AND MINORITY STUDENTS

Comparing the coefficients of the Part I and Part II equations for majority and minority students gave us the following list of variables that may have separate minority and majority student coefficients: school dummy variables, selectivity index, science

MCAT, quantitative MCAT, and verbal MCAT. We fitted regression equations to Part I and Part II scores with all of these variables (as well as several subsets of them) having separate minority and majority student coefficients. This led us to decide that the following specifications were satisfactory for predicting minority and majority student scores from the same equation.

1. The regression equation for Part I scores had a single coefficient for each variable except for (a) separate school dummy variables; (b) a dummy variable indicating a minority student; and (c) separate minority and majority student coefficients for science MCAT, quantitative MCAT, and verbal MCAT.

2. The regression equation for Part II scores had a single coefficient for each variable except for (a) separate school dummy variables, (b) a dummy variable indicating a minority student, and (c) a separate minority and majority student coefficient for science MCAT.

These equations are presented and discussed in Sec. IV. There is a greater difference between the predicted Part I scores for minority and majority students than between the predicted Part II scores. The use of separate minority and majority student quantitative MCAT and verbal MCAT variables in predicing Part I scores, but not in predicting Part II scores, may reflect a "homogenizing" effect of medical school. That is, since Part II is taken about two years later than Part I, the effect of these preadmission attributes may be indistinguishable for majority and minority students after they have been in medical school several years.

Footnotes--Appendix B

[1]See N. Draper and H. Smith, *Applied Regression Analysis*, Wiley, New York, 1966, Ch. 3, for an excellent description of the properties and rationale behind residual analysis.

[2]See Sec. II for a description of Barron's selectivity index.

[3]Rolph, Williams, and Lee (1977).

[4]See J. Tukey, "Discussion, Emphasizing the Connection Between Analysis of Variance and Spectrum Analysis," *Technometrics*, Vol. 3, No. 2, May 1961, for a discussion of this "parsimony in parameters" principle.

[5]Additional variables did not reduce the unexplained variance by a statistically significant amount.

Appendix C

PREDICTION ERRORS FOR COMPARING MINORITY AND MAJORITY STUDENT PERFORMANCE

In Section IV we use our regression equations to characterize the relative performances of minority and majority students on the NBME exams. In this appendix, we describe the formulas used for the prediction errors that are presented in Sec. IV.

Write the minority regression model for a particular NBME score as:

$$Y_i = \alpha_0 + \sum_{j=1}^{k} \alpha_j x_{ij} + \varepsilon_i \ , \qquad (C.1)$$

where Y_i is minority student i's score, $(x_{i1}, \ldots, x_{ik})$ are student i's values of the k independent variables, $\alpha_0, \alpha_1, \ldots, \alpha_k$ are the true regression coefficients, and ε_i is the difference between minority student i's actual and expected scores. That is, the expected score of a minority student with characteristics $\underset{\sim}{x}_i' = (x_{i1}, \ldots, x_{ik})$.

$$\mu_Y(\underset{\sim}{x}_i) = E(Y_i|\underset{\sim}{x}_i) = \alpha_0 + \sum_{j=1}^{k} \alpha_j x_{ij} \ . \qquad (C.2)$$

Let $(a_0, a_1, \ldots, a_k)$ be the least squares estimates of the true coefficients $(\alpha_0, \alpha_1, \ldots, \alpha_k)$. Then our estimate of the *expected performance of a minority student with characteristics* $\underset{\sim}{x}$ *is*:

$$\hat{Y}(\underset{\sim}{x}_i) = a_0 + \sum_{j=1}^{k} a_j x_{ij} \ . \qquad (C.3)$$

Similarly, the majority regression model is

$$Z_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + \varepsilon_i' \qquad (C.4)$$

$$\mu_Z(\underset{\sim}{x}_i) = E(Z_i|\underset{\sim}{x}_i) = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} \ , \qquad (C.5)$$

with $b_0, b_1, \ldots, b_k$ being the estimated least square estimates of $(\beta_0, \beta_1, \ldots, \beta_k)$, and

$$\hat{z}_1(\underset{\sim}{x}_1) = b_0 + \sum_{j=1}^{k} b_j x_{ij} \ . \tag{C.6}$$

Now both $\hat{Y}$ and $\hat{Z}$ are estimates of μ_Y and μ_Z so that they have probability distributions. In Sec. VI we make probabilistic statements about μ_Y and μ_Z as well as their difference, $\mu_Y - \mu_Z$. Under the usual assumptions as given below, the distribution of $\hat{Y}$ is

$$\hat{Y}(\underset{\sim}{x}) \sim N[\mu_Y, \sigma^2(\hat{Y})] \ , \tag{C.7}$$

where "~ N(A, B)" means "has a normal probability distribution with mean A and variance B." The required assumptions are that the distribution of the errors is

$$\varepsilon_i \sim N(0, \sigma_Y^2) \ ; \tag{C.8}$$

and, as a result, the variance of the prediction $\hat{Y}(\underset{\sim}{x})$ is

$$\sigma^2[\hat{Y}(\underset{\sim}{x})] = \underset{\sim}{x}'(X' \ X)^{-1} \ \underset{\sim}{x} \ \sigma_Y^2 \ , \tag{C.9}$$

where $\underset{\sim}{x}$ is vector of characteristics where the prediction is being made and X is the n x k matrix of all the independent variables; i.e.,

$$X = \begin{pmatrix} \underset{\sim}{x}_1' \\ \vdots \\ \underset{\sim}{x}_n' \end{pmatrix} \ .$$

In Sec. IV, the quantity $\sigma^2[\hat{Y}(\underset{\sim}{x})]$ is generally referred to as the *prediction variance of the regression of Y at the point* $\underset{\sim}{x}$. Those readers unfamiliar with matrix notation can best understand the idea using a regression equation with one explanatory variable: $y = \alpha_0 + \alpha_1 x +$ error. When this equation is fitted by least squares

to data points $(x_1, y_1), \ldots, (x_n, y_n)$, the prediction variance for a new point x_0 is given by

$$\sigma_Y^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\Sigma(x_i - \bar{x})^2} \right),$$

where σ_Y^2 is the variance of the error. Note that the further x_0 is from $\bar{x}$ (the average value of x_i), the larger the prediction error.

The analogous statements hold for $\sigma^2(\hat{Z})$ in the majority student case. These formulas are used in Sec. VI, with σ_Y^2 and σ_Z^2 replaced by the estimates:

$$\hat{\sigma}_Y^2 = \frac{1}{n-k-1} \sum_{i=1}^{n} [Y_i - \hat{Y}(\underset{\sim}{x}_i)]^2 \qquad (C.10)$$

The above formulas allow confidence statements and significance tests about expected minority or majority student performance given the students' characteristics--i.e., $\mu_Y(\underset{\sim}{x})$ and $\mu_Z(\underset{\sim}{x})$. The difference $\mu_Y(\underset{\sim}{x}) - \mu_Z(\underset{\sim}{x})$ between expected performance of two comparable minority and majority students is also of great interest. The distribution of the corresponding estimate is

$$D(\underset{\sim}{x}) = \hat{Y}(\underset{\sim}{x}) - \hat{Z}(\underset{\sim}{x}) \sim N[\mu_Y(\underset{\sim}{x}) - \mu_Z(\underset{\sim}{x}), \sigma^2(\hat{Y}) + \sigma^2(\hat{Z})]. \qquad (C.11)$$

Equation (C.11) is used to make inferences about the *expected difference* in performance between a minority and a majority student, both with characteristics $\underset{\sim}{x}$. The dependence of this distribution on $\underset{\sim}{x}$ allows us to characterize where *disparties between average minority and average majority student performance occur.*

Besides making inferences about *average performance*--$\hat{Y}$ or $\hat{Z}$, we are interested in the probability distribution of the difference in performance between a randomly selected minority student $Y(\underset{\sim}{x})$ with characteristics $\underset{\sim}{x}$ and a randomly selected majority student $Z(\underset{\sim}{x})$ with those same characteristics. It is straightforward to see that the distribution of the difference is

$$Y(\underset{\sim}{x}) - Z(\underset{\sim}{x}) \sim N[\mu_Y(\underset{\sim}{x}) - \mu_Z(\underset{\sim}{x}),\ \sigma_Y^2 + \sigma_Z^2] \ . \qquad \text{(C.12)}$$

Since we do not *know* μ_Y and μ_Z, we will need to know that

$$\frac{Y(\underset{\sim}{x}) - Z(\underset{\sim}{x}) - [\hat{Y}(\underset{\sim}{x}) - \hat{Z}(\underset{\sim}{x})]}{\sqrt{\sigma_Y^2 + \sigma^2[\hat{Y}(\underset{\sim}{x})] + \sigma_Z^2 + \sigma^2[\hat{Z}(\underset{\sim}{x})]}} \qquad \text{(C.13)}$$

has a standard normal distribution. The distribution of (C.13) allows us to estimate the probability that a randomly selected majority student's performance will exceed a comparable (same $\underset{\sim}{x}$) randomly selected minority student's performance. Specifically, write the numerator to Eq. (C.13) as

$$\{Y(\underset{\sim}{x}) - Z(\underset{\sim}{x}) - [\mu_Y(\underset{\sim}{x}) - \mu_Z(\underset{\sim}{x})]\} - \{\hat{Y}(\underset{\sim}{x}) - \hat{Z}(\underset{\sim}{x}) - [\mu_Y(\underset{\sim}{x}) - \mu_Z(\underset{\sim}{x})]\} \ .$$

The first term in braces has a normal distribution with mean 0 and variance $\sigma_Y^2 + \sigma_Z^2$, and the second term in braces is normally distributed with mean 0 and variance $\sigma^2[\hat{Y}(x)] + \sigma^2[\hat{Z}(x)]$. Since these two terms are independent, (C.13) is standard normal, and the probability that $Y(\underset{\sim}{x}) > Z(\underset{\sim}{x})$ is

$$\Phi\left\{\frac{\mu_Y(\underset{\sim}{x}) - \mu_Z(\underset{\sim}{x})}{\sqrt{\sigma_Y^2 + \sigma_Z^2 + \sigma^2[\hat{Y}(\underset{\sim}{x})] + \sigma^2[\hat{Z}(\underset{\sim}{x})]}}\right\} ,$$

where Φ is the distribution function of a standard normal random variable. The probabilities in Tables 7 and 9 are computed in this way.

Footnote--Appendix C

[1]Strictly speaking, the resulting distribution of $\hat{\mu}_M - \hat{\mu}_M / \hat{\sigma}_M$ is students-t with n-k-1 degrees of freedom. See N. Draper and H. Smith, *Applied Regression Analysis*, Wiley, New York, 1966.

"END OF DOCUMENT"